Integrated Feature Analysis for Prostate Tissue Characterization Using TRUS Images

by

Samar Samir Mohamed

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis.

This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Signature

Abstract

The Prostate is a male gland that is located around the urethra. Prostate Cancer is the second most diagnosed malignancy in men over the age of fifty. Typically, prostate cancer is diagnosed from clinical data, medical images, and biopsy.

Computer Aided Diagnosis (CAD) was introduced to help in the diagnosis in order to assist in the biopsy operations. Usually, CAD is carried out utilizing either the clinical data, using data mining techniques, or using features extracted from either TransRectal UltraSound (TRUS) images or the Radio Frequency (RF) signals.

The challenge is that TRUS images' quality is usually poor compared to either Magnetic Resonance Imaging (MRI) or the Computed Tomography (CT). On the other hand, ultrasound imaging is more convenient because of its simple instrumentation and mobility capability compared to either CT or MRI. Moreover, TRUS is far less expensive and does not need certain settings compared to either MRI or CT. Accordingly; the main motivation of this research is to enhance the outcome of TRUS images by extracting as much information as possible from it. The main objective of this research is to implement a powerful noninvasive CAD tool that integrates all the possible information gathered from the TRUS images in order to mimic the expert radiologist opinion and even go beyond his visual system capabilities, a process that will in turn assist the biopsy operation. In this sense, looking deep in the TRUS images by getting some mathematical measures that characterize the image and are not visible by the radiologist is required to achieve the task of cancer recognition.

This thesis presents several comprehensive algorithms for integrated feature analysis systems for the purpose of prostate tissue classification. The proposed algorithm is composed of several stages, which are: First, the regions that are highly suspicious are selected using the proposed Gabor filter based ROI identification algorithm.

Second, the selected regions are further examined by constructing different novel as well as typical feature sets. The novel constructed feature sets are composed of statistical feature sets, spectral feature sets and model based feature sets.

Next, the constructed features were further analyzed by selecting the best feature subset that identifies the cancereus regions. This task is achieved by proposing different dimensionality reduction methods which can be categorized into: Classifier dependent feature selection (Mutual Information based feature selection), classifier independent feature selection, which is based mainly on tailoring the Artificial life optimization techniques to fit the feature selection problem and Feature Extraction, which transforms the data to a new lower dimension space without any degradation in the information and with no correlation among the transformed lower dimensional features.

Finally, the last proposed fragment in this thesis is the Spectral Clustering algorithm, which is applied to the TRUS images. Spectral Clustering is a novel fast algorithm that can be used in order to obtain a fast initial estimate of the cancer regions. Moreover, it can be used to support the decision obtained by the proposed cancer recognition algorithm. This decision support process is crucial at this stage as the gold standards used in obtaining the results shown in this thesis is mainly the radiologist's

markings on the TRUS images. This gold standards is not considered as credible since the radiologist's best accuracy is approximately 65 %.

In conclusion, this thesis introduces different novel complete algorithms for automatic cancerous regions detection in the prostate gland utilizing TRUS images. These proposed algorithms complement each other in which the results obtained using either of the proposed algorithms support each other by resulting in the same classification accuracy, sensitivity and specificity. This result proves the remarkable quality of the constructed features as well as the superiority of the introduced feature selection and feature extraction methods to detect cancerous regions in the prostate gland.

Acknowledgements

With genuine humility, I am obediently thankful to Almighty Allah, the only one God, praise and glory be to Him, for all His uncountable bounties and guidance, without which, this work would have never become a reality.

I would like to express my sincere thanks to my supervisor, Professor Magdy Salama, for his constant support, guidance, fruitful discussions and encouragement throughout the course of this work. He was always there when I needed him.

Words fall short of expressing my love, appreciation and gratitude for my dear parents for being always there for us. Without their love, encouragement and support this work would have never been possible.

Special thanks to my husband, Ehab, for his love, understanding, and support throughout the course of this work, and to my beloved daughters Menna, Hajar, Reem and Farah, whose innocent smiles make all the difference to my world.

I would like to thank my colleagues Dr. Amr Youssef and Khaled Hammouda for their useful discussions. I would also like to thank Dr. Kamilia Rizkalla, Prof. Mohamed Kamel and Prof. George Freeman for the feedback they gave me during my research years. I would like also to thank my external examiner Prof. Khaled Kamel for his interest in my work. Many thanks go to Professor Maher Ahmed for showing interest in my research and for being in my exam committee.

My thanks go to the members of Electrical and Computer Engineering department, especially Wendy Boles for her endless support and help in solving my problems.

Moreover, many thanks go to Prof. Ed Jernigan and Prof. Andrew Wong for the useful graduate courses that had lots of influence on my research.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction and Objectives

1.1 Introduction

The Prostate is a male gland that is located around the urethra. The Prostate is considered as four glandular structures, which are the Anterior zone, the Peripheral zone, the Central zone and the Transition zone [1, 2]

Prostate Cancer is the second most diagnosed malignancy in men over the age of fifty [3]. Symptoms due to carcinoma of the Prostate are generally absent until extensive local growth or metastases develop. Yet, risk of metastasis dramatically increases once the tumor has extended beyond the prostate [3]. Therefore, when diagnosed at an early stage, the disease is curable. Even at later stages, treatment can be effective. Nevertheless, treatment options vary depending on the cancer extent. Traditionally, prostate cancer screening is achieved from clinical data, medical images, and biopsy [3]. Recently, Computer Aided Diagnosis (CAD) was incorporated to help in the screening stage in order to assist in the biopsy operations. CAD was carried out utilizing the clinical data, either using data mining techniques, or using features extracted from TransRectal UltraSound (TRUS) images.

TRUS images' quality is usually poor and noisy compared to Magnetic Resonance Imaging (MRI). On the other hand, ultrasound imaging is more convenient because of its simple instrumentation and mobility capability compared to both CT and MRI. Moreover, TRUS is far less expensive and does not need special settings compared to either MRI or CT. Therefore, the main objective of this research is to implement a powerful noninvasive CAD tool that integrates all the possible information gathered from the TRUS images in order to mimic the expert radiologist opinion and even go beyond his visual system capabilities, a process that will in turn assist the biopsy operation. This technique should lead to the improvement of prostate cancer diagnosis utilizing ultrasound images and ultimately make the outcome of this method comparable to that of the other image modalities. Moreover, the proposed technique is expected to be crucial in guiding the biopsy operation by providing the high probable cancer regions.

1.2 Overview of Prostate Cancer Diagnosis Methods

Digital Rectal Exam (DRE) is a clinical test usually conducted in the doctor's office where the doctor feels the prostate gland through the rectum boundary. Any abnormality in either the texture or volume of the prostate might indicate prostate cancer. DRE might easily miss tumors that are small or deeply located in the gland. DRE was reported to miss 44–59% of cancers in [4], and can only access tumors in specific areas of the prostate. Prostate-specific antigen (PSA) is a protein produced by the prostate cells; its concentration is measured in the blood stream where an elevated PSA level might indicate prostate cancer. PSA level and DRE are considered the typical assessments for screening individuals for prostate cancer [5].

The combination of these two screening techniques (DRE and PSA) has improved the diagnostic path by allowing for an earlier detection of tumors [6]. However, PSA has limited specificity in that it is not a tumor marker, but its value can increase due to different effects such as Prostate volume increase, BPH and Prostatitis. Abnormal PSA values, specifically those in the 4 to 10 ng/mL range, can occur in men with benign Prostate conditions such as benign prostate hyperplasia and prostatitis. Only 25% of individuals that undergo biopsies are actually positive for prostate cancer. Conversely, low PSA findings are not always a true indicator of a healthy prostate. A recent study showed that 15% of individuals with PSA levels below four ng/mL have prostate cancer as detected by biopsy [7].

Another screening way is the determination of the prostate volume. The prostate volume is a crucial parameter in the diagnosis and management of both benign and malignant prostatic diseases. However, prostate volume increases with several factors such as age and BPH and therefore, is not considered an accurate screening measure [8]. Yet, it is still considered among the attributes that determine the necessity of performing biopsy operation. The Prostate volume is usually estimated by using CT imaging system, MRI imaging or TRUS imaging system. It was shown in [9] that the volume obtained using TRUS is more accurate than the volume obtained using CT. In a comparison between the TRUS imaging, and the DRE, the DRE obtained far less staging ability than the TRUS [10].

From the previous discussion as well as the affordable price and the simplicity, it is clear that using the TRUS imaging modality is the most favorable. Moreover, the internal architecture of the prostate can be detailed by TRUS, and the procedure allows an accurate measurement of the prostate volume. Therefore, TRUS has become the most frequently used prostate imaging modality and is becoming a recommended step for prostate cancer screening. Furthermore, as cancerous tissues need more blood supply than non-cancerous tissues, it builds its own blood supply, which makes these regions different when imaged in the procedure of prostate cancer diagnosis [11]. Prostate carcinoma typically has a characteristic hypo-echoic pattern, which is sometimes distinct from the normal echo pattern of the peripheral portion of the prostate. However, TRUS is still lacking sufficient sensitivity and specificity to support its use for routine screening for prostate cancer. In men with an abnormality of either DRE of the prostate or serum PSA, TRUS is useful for directing prostate biopsy. This is usually achieved with the aid of an expert radiologist that marks the suspicious regions in the TRUS image [12]. During the biopsy procedure, an ultrasound probe is used to guide the placement of the biopsy needle. The biopsy needle is used to obtain tissue samples from different areas of the prostate. The standard procedure is to obtain six samples from uniformly distributed locations in the gland. The pathologist usually examines these tissues and determines the presence of cancer. Unfortunately, some of the cancer cases are missed in this procedure due to the way tissue samples are collected during the biopsy.

1.3 Motivations

Biopsies are usually taken from the suspected lesions identified by DRE or TRUS. Otherwise, when no suspicious region in the prostate is present, but the PSA-value is higher than 10 ng/ml a sextant random biopsy is performed [8]. In some situations, the biopsy results are false negative due to the

inaccurate biopsy location [13, 14]. This highlights the importance of aiding the biopsy operation. Therefore, estimating the accurate locations for the biopsy is a crucial phase for prostate cancer diagnosis. TRUS imaging can play a significant rule in prostate cancer diagnosis by providing precise information that can lead to accurate biopsy operation.

Identifying the suspicious regions in the prostate from TRUS is usually achieved by the aid of expert radiologist, where experienced radiologists have higher predictive rate than inexperienced radiologists. Therefore, there is a great need to augment the knowledge of the newer radiologists with the information of the more experienced ones.

Texture features such as statistical features, spectral features, fractal dimension and Laws Energy Measures (LEMs) features are usually used individually or combined in small groups for designing a CAD system for tissue classification involving ultrasound images. These methods did not yield the required accuracy in case of prostate cancer. One of the main reasons for these inaccurate results is the use of the raw unanalyzed features, where all the constructed features were used and no selection was ever performed to identify the most salient features that properly represent the cancerous tissues. Another reason for the poor results is the way images were studied. The first method was analyzing the images as a whole where it was divided to small rectangles and each rectangle is considered a region. This method yields poor accuracy as cancerous and non-cancerous regions might be combined in any of those rectangles and the constructed features are distorted. The other technique of image interpretation was the analysis of only the regions marked by the radiologist. This method is not recommended as it depends mainly on the conducting radiologist and it might carry on the radiologist's mistake if he misidentifies or looses any region.

1.4 Objectives

The main goal of this work is to assist the radiologist in the decision making process and to augment his findings in order to accurately choose the biopsy locations. This will lead ultimately to get the samples from the most suspicious regions, which in turn will decrease the false negative rate. This is achieved mainly by studying and analyzing different features of the TRUS images and by integrating the information from the expert medical doctors into the feature analysis algorithm. A complete CAD system is implemented in this work in order to analyze all the possible texture features. Moreover, some novel features are also proposed and utilized to enhance the accuracy of the system.

Another objective of this research is to extend the ultrasound prostate imaging system capabilities as a vital diagnostic tool. In this research, The TRUS images are utilized to develop a set of tools with novel functionality, which enables the user to extract useful information from TRUS images about the prostate anatomic structure. The extraction and the classification of this information are based on the knowledge of the expertise in the field. In this research, feature extraction, analysis and selection were tackled.

In order to create significant visualization and identify tumors in TRUS images, internal segmentation which is intended to subdivide the gland into its parts is used in this work. It allows for the choice of some Regions Of Interest (ROIs) for further analysis. Manual and semi-automatic segmentation tools

were already applied in practice in different fields. Several segmentation algorithms that segment the prostate image from the background (boundary detection) were already proposed in literature, yet prostate internal segmentation (ROI identification) was not tackled in literature, to the best of the author's knowledge. The main goals of this work are:

- To develop an efficient, automatic and reliable technique that identifies the ROIs in the TRUS images (internal segmentation). The superiority of this technique lies in the incorporation of the medical expertise knowledge with the multi-resolution analysis in order to reduce the time and effort necessary for human interaction.
    - To extract useful features from the segmented structure (ROI). These features are based on the statistical, the spectral and the model based foundations of the segmented ROIs. In this research, those sets of features are tested and integrated to fully understand and gather information from the image. This analysis is expected to define the location of the tumor inside the volume of the segmented structure.
    - To select the most relevant features and reduce the redundancy utilizing different feature selection algorithms and different dimensionality reduction techniques.
    - To apply a simple classifier in each stage in order to judge this specific feature set. The same classifier is applied after each feature selection algorithm in order to judge the new feature subset.
    - To support the decision obtained using the CAD system by another segmentation method that looks at the problem from a different point of view. Spectral clustering is a recently proposed image segmentation algorithm that treats the image as a graph and looks for the optimum graph cut. Spectral clustering was applied in this work for the first time with TRUS image segmentation. The Spectral clustering is used in this work first in order to segment the prostate from the background (boundary detection), a task that is performed mainly to prove the credibility of the spectral clustering method by comparing its' segmentation results to the gold standards. Once the Spectral clustering method proves its accuracy, it is then used for internal segmentation to identify different regions in the gland.

1.5 Proposed Algorithm

With trans-rectal ultrasound, there is the potential to categorize tissues belonging to different pathological groups. This can ultimately aid in the understanding of individual lesions as well as making a significant contribution to treatment choice and management of cancer patients. However, this visual diagnosis is in general highly subjective due to observer errors. To overcome the shortcomings inherent in the visual assessment of ultrasound images, texture analysis is used to assess abnormalities in the ultrasound image. Texture is perceived as homogeneous visual patterns representing the composition of the surface being imaged. Since every tissue sub-group has its own texture, verified from histological analysis, then, texture can be used as a way of characterizing the tissue subgroup.

The ultimate objective of this work is to determine a reliable automated procedure based on the integration between all valuable TRUS image textural features for assessing prostate cancer. The TRUS images corresponding to those suspicious cases (due to either DRE or PSA abnormality) were analyzed by using multi-resolution texture analysis in order to determine the ROIs. Furthermore, second and higher-order statistics were then implemented to extract useful statistical features from the pre-determined ROIs. In addition, Spectral features were constructed using different techniques as another piece of information that helps in prostate cancer recognition. In a further attempt to enhance the aforementioned extraction techniques, fractal texture analysis was used to obtain textural information on regions of interest. Fractals proved to be useful in describing the texture of these areas by a single measure. This measure, the fractal dimension, describes the degree of irregularity in the surface texture. Moreover, the Laws Energy Measures are also considered as valuable features that were used for the first time for TRUS image analysis. In order to reduce the computational complexity, and remove redundant information, different feature selection methods were implemented. Moreover, in order to support the decision using a different foundation, Spectral clustering was also used not only to segment the gland from the background (boundary detection) in order to test the Spectral Clustering credibility, but also to segment the regions of the gland. The steps of the proposed research are summarized in Figure 1-1.

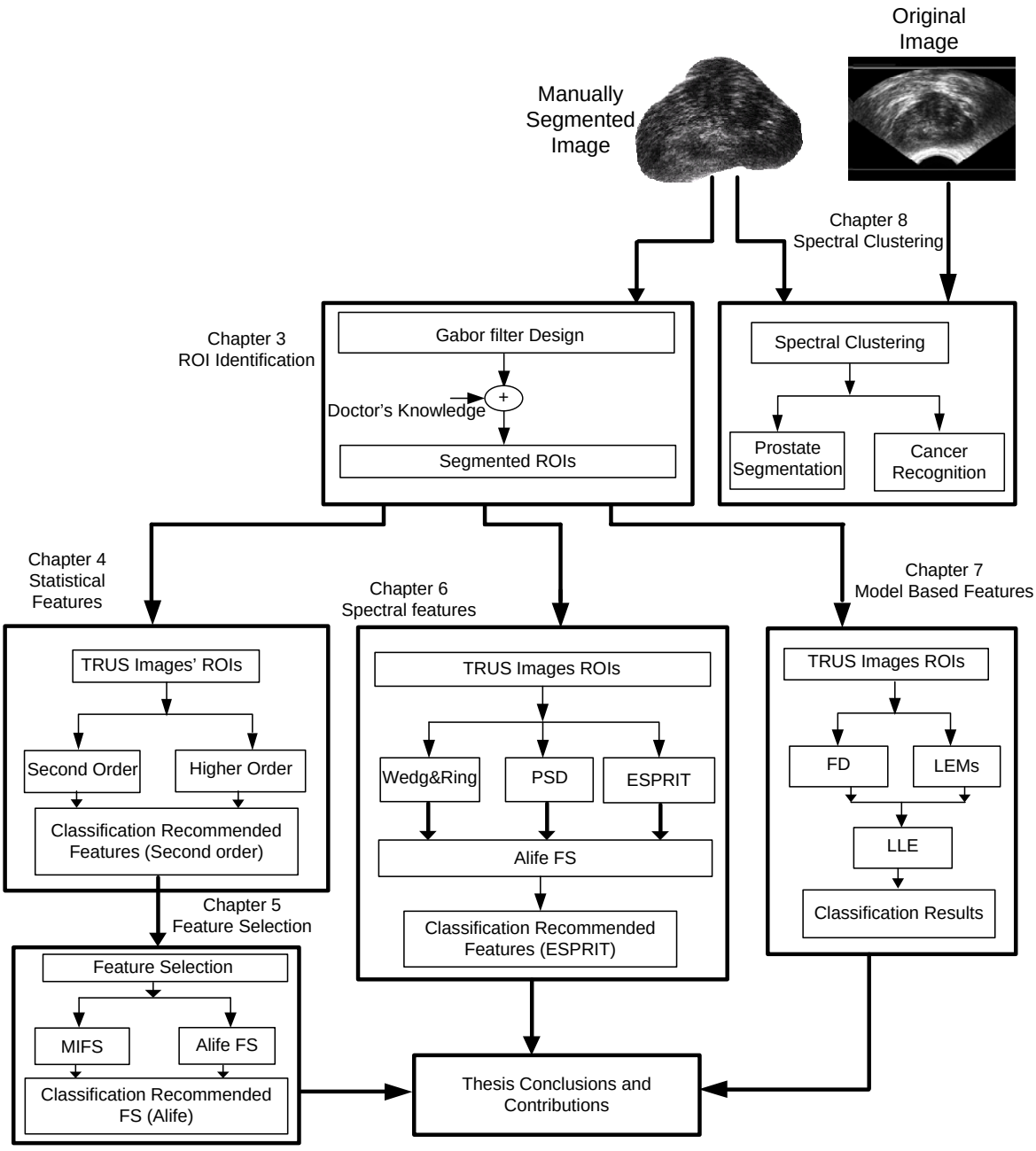


Figure 1-1 Proposed Algorithm Summary

The proposed algorithm is composed of several stages, which are:

ROI Identification: where the highly suspicious cancerous regions are identified for further analysis this stage is shown in Figure 1-2.

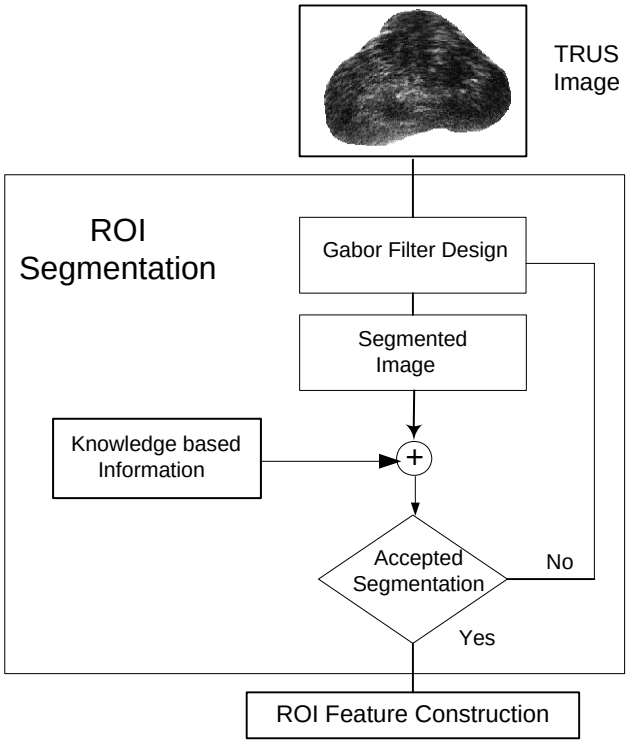


Figure 1-2 ROI Segmentation Algorithm

Feature Construction: where several features are constructed from the identified ROIs. These features are mainly:

Statistical features: Grey Level Difference Matrix (GLDM), Grey Level Difference Vector (GLDV), Neighborhood Grey Tone Difference Matrix (NGTDM) and Grey Tone Run Length Matrix (GTRLM).

*Spectral feature*s: Wedge and ring filters, parametric Power Spectrum Density (PSD) estimation methods, non- parametric PSD estimation methods, Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT).

The Fractal Dimension (FD) for each identified ROI is estimated using three different methods; moreover, the *Laws Energy Measures* (LEMs) for each ROI are also constructed.

.Those methods are highlighted in Figure 1-3.

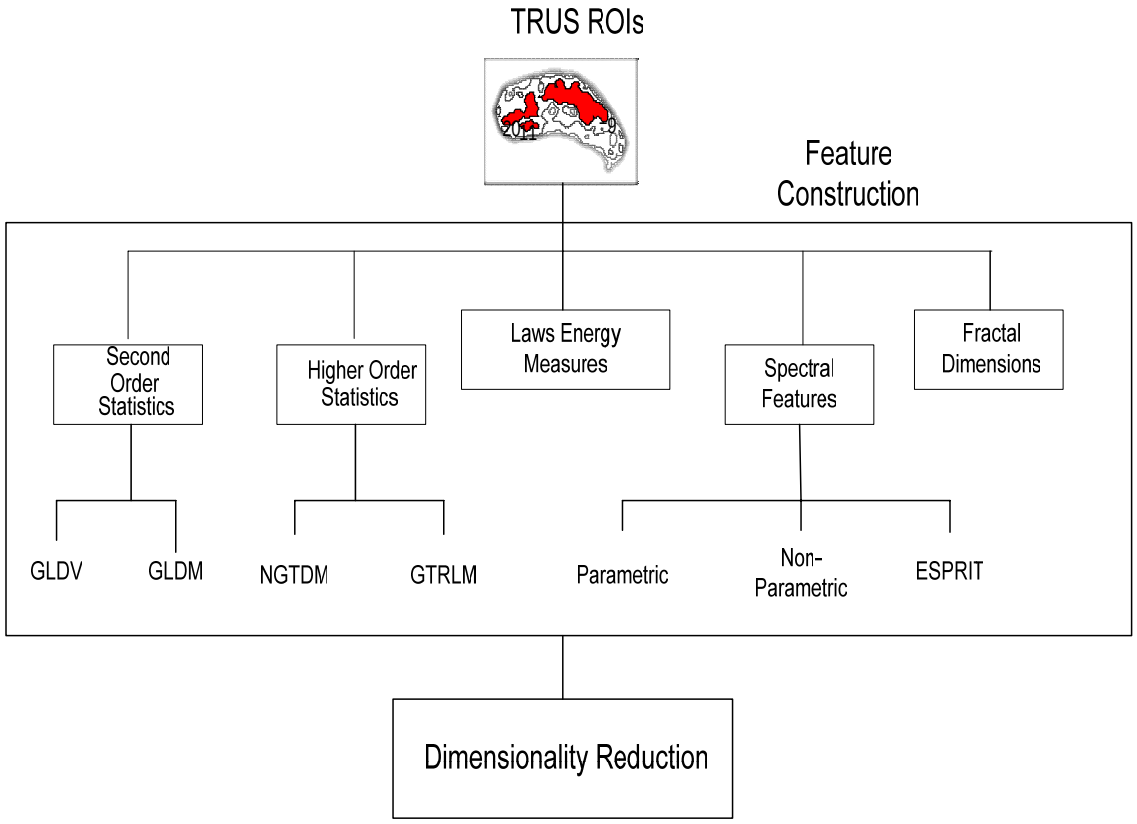


Figure 1-3 Feature Construction Methods

Dimensionality reduction: this step is achieved in this work by either feature selection or feature extraction. By using the feature selection avenue the most salient and informative features are selected among each of the constructed feature sets. Further, by using the feature extraction avenue a new smaller less correlated feature set, yet carrying the important information for classification is extracted from the available feature sets. Feature selection can be categorized mainly to classifier dependent and classifier independent feature selection methods. Both approaches are adopted and customized in this work, where *Mutual Information Feature Selection* (MIFS) is used as a classifier independent feature selection method. Moreover, several *Artificial life (Alife)* optimization techniques were introduced in this work for the first time in the field of TRUS image feature selection problem. Alife optimization techniques are modified and applied as classifier dependent feature selection methods to fit the TRUS image feature selection. Moreover, the *Locally Linear Embedding* (LLE) is also introduced in this work for TRUS image features dimensionality reduction. Different feature selection and dimensionality reduction methods are illustrated in Figure 1-4.

The final stage in the feature analysis is testing the selected features using the SVMs classifier. SVMs is well known for its ability to deal with non-linearly separable data, which is the case of TRUS image features. Because the different parameters have a highly nonlinear interdependence which is

discussed in section 5.4.1, only a nonlinear model is able to classify the incidences and, thus, lead to reliable results [15].

Spectral Clustering: the original TRUS image was segmented using the Spectral Clustering algorithm that depends mainly on representing the image by a weighted graph and finding the best cut for that graph. The segmented gland was also segmented using the Spectral Clustering to obtain different regions in the gland, where cancerous regions can be identified.

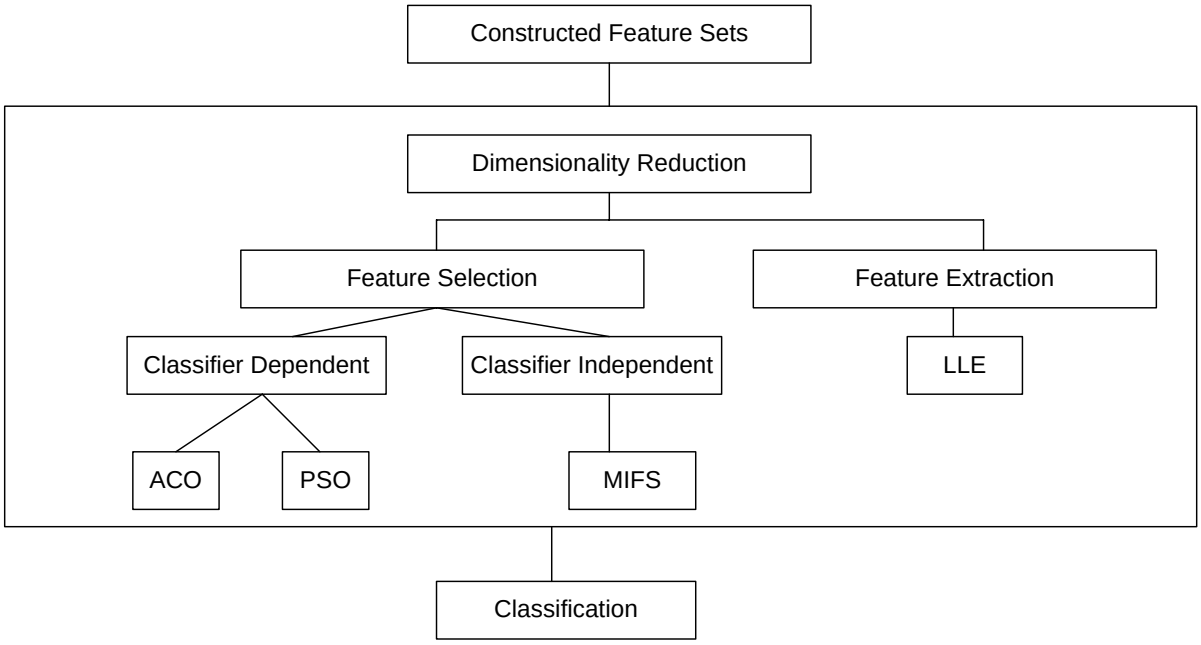


Figure 1-4 Dimensionality Reduction Methods

1.6 Organization of the Thesis

Following the background (Chapter 2), the proposed ROI identification algorithm using Gabor multi-resolution analysis combined with the expertise knowledge is explained and the identified ROIs are shown in Chapter 3.

In Chapter 4 the different Statistical feature construction methods including second and higher order statistics that are implemented and used for ROI statistical feature construction. These features are also tested using simple classifiers in the same chapter.

Chapter 5 focuses on different feature selection methods applied for the statistical features constructed in Chapter 4 and the most appropriate feature selection method is chosen to be applied throughout the thesis.

In Chapter 6 different Spectral features are constructed from each identified ROI and the most discriminatory spectral features are then selected using the feature selection algorithms recommended from Chapter 5.

Chapter 7 shows the construction of different textural features which are mainly the FD and LEMs. The constructed features are then subjected to a feature extraction method (LLE) that transforms the constructed features to a different domain where they can be easily separable.

Chapter 8 summarizes the Spectral Clustering algorithm and shows the obtained segmentation results for both the whole gland as well as the gland regions segmentation.

A summary and discussion is finally shown in Chapter 9 as well as the recommendation for future work.

Chapter 2

Literature Review

2.1 Preamble

The prostate is a complex organ consisting of acinar, stromal, and muscular elements. Anatomically it is in the shape of a compressed inverted cone gland it is situated in the pelvis and lies in front of the ampulla of the rectum. Its upper end is continuous with the neck of the bladder, and its apex rests on the superior aspect of the fascia of the urogenital diaphragm. The Urethra (the tube that transports urine and sperm out of the body) runs through the prostate from base to apex as shown in Figure 2-1 and Figure 2-2. The prostate secretes seminal fluid, a milky substance that combines with sperm produced in the testicles to form semen. Prostate tissue produces prostate specific antigen and prostatic acid phosphatase, an enzyme found in seminal fluid. The medical information in this chapter is mainly inspired from [16, 17, 18, 19, 20]

2.2 Prostate Anatomy

The prostate is located directly beneath the bladder and in front of the rectum. It has Anterior, Transition, and Central and Peripheral zone as shown in figure 1. 60 to 70 % of the cancer occurs in the Peripheral zone, 10 to 20% occurs in the Central zone and the Benign mainly occurs in the Transition zone. Because the upper portion of the urethra passes through the prostate, if the gland becomes enlarged, it can obstruct the passage of fluids through the urethra. The normal prostate measures 3-4 cm in its widest portion, 4-6 cm in its length and 2-3 cm in its thickness. The prostate size increases variably with malignancy. Structurally it is composed of fibro muscular tissue and glandular cells. The fibro muscular component is present mostly anteriorly, while the glandular element is mostly in the peripheral and transition aspects of the gland.

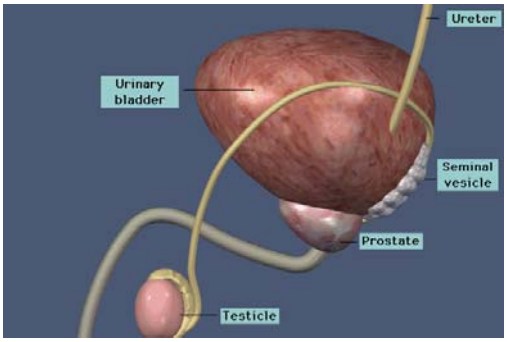


Figure 2-1 Location of the Prostate Gland

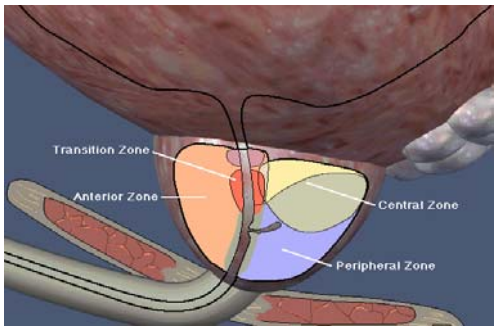


Figure 2-2 Zones of the Prostate Gland

2.3 Benign Diseases of the Prostate

Prostate problems are common in men age 50 and older. Doctors who are experts in diseases of the urinary tract (urologists) diagnose and treat prostate problems. There are many different kinds of prostate problems. Many do not involve cancer, but some do.
An enlarged prostate eventually develops in approximately 80 percent of all men. Benign diseases are non-cancerous. They can be uncomfortable and inconvenient but are not life threatening, and often can be treated with drugs or surgery. The two main benign diseases of the prostate are Prostatitis and Benign Prostatic Hyperplasia (BPH).

2.3.1 BPH

Benign prostatic hyperplasia (BPH) is not only a case of too many prostate cells. As a result of these differences, treatment varies in each case. There is no cure for BPH and once prostate growth starts, it often continues, unless medical therapy is started. The prostate grows in two different ways. In one type of growth, cells multiply around the urethra and squeeze it. The second type of growth is middle-lobe prostate growth in which cells grow into the urethra and the bladder outlet area. This type of growth typically requires surgery. Most men over the age of 50 begin to experience symptoms of the condition.

2.3.1.1 Diagnosis

A physical examination, patient history, and evaluation of symptoms provide the basis for a diagnosis of BPH. The physical examination includes a digital rectal examination (DRE).

2.3.1.1.1 Digital rectal Examination (DRE)

DRE is an uncomfortable procedure that typically takes less than a minute to perform. The doctor inserts a lubricated, gloved finger into the patient's rectum to feel the surface of the prostate gland through the rectal wall to assess its size, shape, and consistency. Healthy prostate tissue is soft, like

the fleshy tissue of the hand where the thumb joins the palm. Malignant tissue is firm, hard, and often asymmetrical or stony, like the bridge of the nose. If the examination reveals the presence of unhealthy tissue, additional tests are performed to determine the nature of the abnormality.

2.3.1.1.2 AUA Symptom Index

The AUA (American Urological Association) Symptom Index is a questionnaire designed to determine the seriousness of a man's urinary problems and to help diagnose BPH. The patient answers seven questions related to common symptoms of benign prostatic hyperplasia. How frequently the patient experiences each symptom is rated on a scale of 1 to 5. These numbers added together provide a score that is used to evaluate the condition. An AUA score of 0 to 7 means the condition is mild; 8 to 19, moderate; and 20 to 35, severe.

2.3.1.1.3 PSA and PAP Tests

Blood tests taken to check the levels of prostate specific antigen (PSA) and prostatic acid phosphatase (PAP) in a patient helps the physician in the prostate cancer screening process.

- **PSA**

It is a specific antigen produced by the cells of the prostate capsule (membrane covering the prostate) and periurethral glands. Patients with benign prostatic hyperplasia (BPH) or prostatitis produce larger amounts of PSA. The PSA level also is determined in part by the size and weight of the prostate.

- **Age Specific PSA**

Evidence suggests that the PSA level increases with age. A PSA up to 2.5 ng/mL for men age 40 – 49 is considered normal, as is 3.5 ng/mL for men age 50–59, 4.5 ng/mL for men age 60–69, and 6.5 ng/mL for men 70 and older.

2.3.1.1.4 Uroflow and Urodynamic Testing

- **Urodynamic**

Urodynamic tests are usually performed in a physician’s office and are used to study the volumes and pressure of urine in the bladder and the flow of urine from the bladder. Additional tests may be conducted if symptoms indicate severe blockage caused by a condition other than BPH.

- **Uroflowmetry**

Uses a device (called a uroflowmeter) to record urine flow, determine how quickly and completely the bladder can be emptied, and evaluate obstruction. With a full bladder, the patient urinates into a device that records the amount of urine, the time it takes for urination, and the rate of urine flow. A reduced flow rate may indicate BPH.

2.3.1.1.5 A Pressure Flow Study

Measures pressure in the bladder during urination and is designed to detect a blockage of flow. It is the most accurate way to evaluate urinary blockage. This test requires the insertion of a catheter through the urethra in the penis and into the bladder. The procedure is uncomfortable and rarely may cause urinary tract infection (UTI).

2.3.1.1.6 Post-Void Residual (PVR)

It measures the amount of urine that remains in the bladder after urination. The patient is asked to urinate immediately prior to the test and the residual urine is determined by ultrasound or catheterization. PVR less than 50 mL generally indicates adequate bladder emptying and measurements of 100 to 200 mL or higher often indicate blockage. The clinical setting may affect the result; therefore, the test is usually repeated.

2.3.2 Prostatitis

Prostatitis, inflammation of the prostate gland, is a common urological condition. More than 2 million men in the United States are diagnosed with prostatitis each year. An infected or inflamed prostate can cause painful urination and ejaculation, and can cause serious complications. Complications of acute bacterial prostatitis (ABP) include: Abscess (collection of pus), acute urinary retention (link), chronic bacterial prostatitis and sepsis (infection in bloodstream)

2.3.2.1 Diagnosis

2.3.2.1.1 DRE

Is used to determine if the prostate gland is tender or enlarged with different texture.

2.3.2.1.2 A 3-part Urinalysis

Is the standard diagnostic tool. Two urine specimens are collected and analyzed, followed by prostate massage and third urine sample that contains prostatic fluid. Urinalysis is used to determine the presence of leukocytes (white blood cells) in the urine. Leukocytes help the body to fight infection; a high number indicates a bacterial infection. A urine culture is used to analyze bacteria.

Nonbacterial prostatitis is diagnosed when tests reveal no bacteria in the urine or prostatic secretions. There is no test to diagnose prostatodynia; it is diagnosed after eliminating other probable causes (e.g., kidney stones, uresthral cancer)

2.4 Prostrate Cancer

Prostate cancer usually arises in the peripheral zone and is often multi centric. The most common histological type is the adenocarcinoma. Adenocarcinoma of the prostate is the clinical term for cancer that begins as a tumor on the prostate gland. Adenocarcinoma of acinar origin comprises 98% of the prostatic cancers. Microscopically the cancer can be well, moderate or poorly differentiated. The well-differentiated cancers show uniform medium sized or small gland formation throughout the tumor. As prostate cancer grows, it may spread to the interior of the gland, to tissues near the prostate, to seminal vesicles (sac-like structures attached to the prostate), and to distant parts of the body (e.g., bones, liver, lungs). Prostate cancer confined to the gland often can be treated successfully.

Some of the risk factors of having prostate cancer are: family history, a diet high in saturated fat, 55 years old and older, exposure to heavy metals, race, smoking and exposure to certain viral infections may also increase the risk for developing prostate cancer.

2.4.1 Signs and Symptoms

Early prostate cancer is not usually discovered during a routine digital rectal examination (DRE). Symptoms of staged prostate cancer often resemble those of BPH. Men experiencing the following symptoms should see their physician immediately for a thorough examination.

- Blood in the urine or semen
- Frequent urination, especially at night
- Inability to urinate
- Nagging pain or stiffness in the back, hips, upper thighs, or pelvis
- Painful ejaculation
- Pain or burning during urination (dysuria)
- Weak or interrupted urinary flow

2.4.2 Staging

Two systems are commonly used for staging prostate cancer: the Jewett-Whitmore system and the TNM (tumor, node, metastases) system.

2.4.2.1 Jewett-Whitmore system

In the Jewett-Whitmore system, prostate cancer is classified first as stage A, B, C, or D. Stages A and B cancers are considered curable [21]. Stages C and D are treatable, but their prognoses are discouraging. A number is then assigned to describe specific conditions within each stage. For example, a tumor classified, as stage B1 is a single cancerous nodule confined to one lobe of the prostate.

2.4.2.1.1 Stage A

Very early and without symptoms, cancer cells confined to the prostate. Illustrated in Figure 2-3

- Tumor not detectable by imaging or clinical exam
- Less than 5% of tissue specimen
- Low grade tumor

A1 Well differentiated and slightly abnormal cancer cells

A2 Moderately or poorly differentiated and abnormal cancer cells in several locations within the prostate

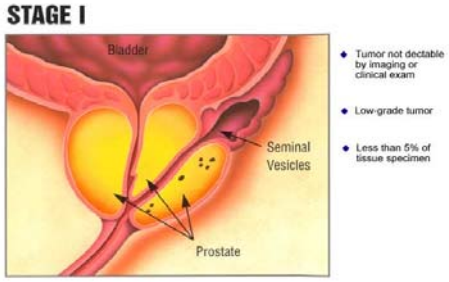


Figure 2-3 Stage A Prostate Cancer

2.4.2.1.2 Stage B

Cancer is confined to the prostate, but palpable (detectable by digital rectal exam) and/or detectable by elevated PSA. It is shown in Figure 2-4.

- Tumor not detectable by imaging or clinical exam (may be found by one or more lobes by needle biopsy).
- Moderate/high grade tumor.
- More than 5% of tissue specimen

B0 Confined to the prostate, non-palpable; PSA elevated

B1 Single cancerous nodule in one lobe of the prostate

B2 Extensive, involvement in one or both prostate lobes

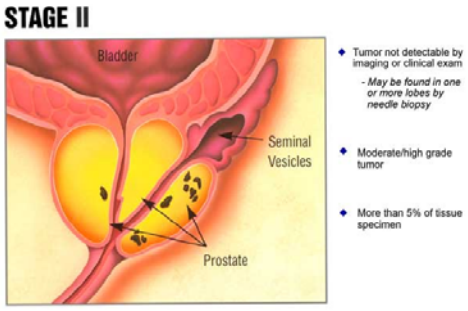


Figure 2-4 Stage B Prostate Cancer

2.4.2.1.3 Stage C

Cancer cells found outside the prostate capsule (membrane covering the prostate); spread confined to surrounding tissues and/or seminal vesicles this stage is shown in Figure 2-5.

- Tumor extends beyond prostate capsule.
- Any grade tumor.

C1 Extends outside the prostate capsule

C2 Bladder or urethral obstruction

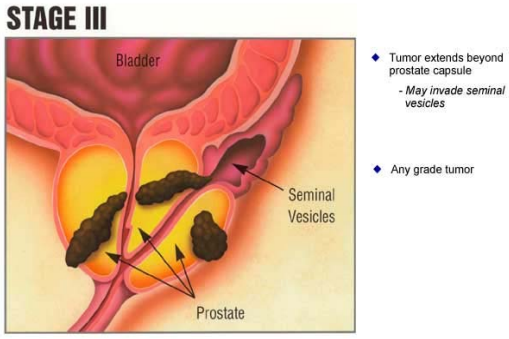


Figure 2-5 Stage C Prostate Cancer

2.4.2.1.4 Stage D

Metastasis (spread) to regional lymph nodes, or to distant bones, organs (e.g., liver, lungs), and/or other tissues, it is illustrated in Figure 2-6.

D0 Metastatic, clinically localized, and showing elevated blood PAP levels

D1 Regional lymph nodes involved

D2 Distant lymph nodes, bones, or organs involve

D3 Metastatic disease after treatment

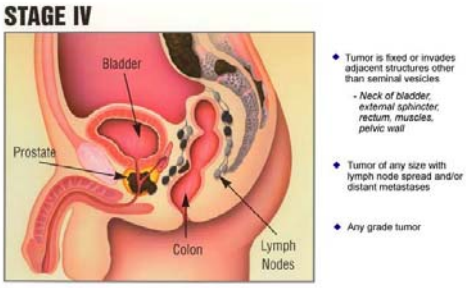


Figure 2-6 Stage D Prostate Cancer

2.4.2.2 TNM System

The TNM (tumor, node, metastases) system stages are similar to those of the Jewett-Whitmore system, but with more specific alphanumeric subcategories [21].

2.4.2.2.1 Primary Tumor (T)

TX Tumor cannot be assessed

T0 No evidence of primary tumor

T1 Clinically not palpable or visible by imaging

T1a Found incidental to other surgery

T1b Found incidental to other surgery

T1c Identified by needle biopsy

T2 Tumor confined within prostate

T2a Involving half a lobe or less of prostate

T2b Involving half a lobe

T2c Involving both lobes

T3 Tumor extends through prostate capsule

T3a Extends through one lobe

T3b Extends through both lobes

T3c Extends into seminal vesicles

T4 Involves structures other than seminal vesicles

T4a Invades bladder neck, external sphincter, or rectum

T4b Invades muscles and/or pelvic wall

2.4.2.2.2 Regional Lymph Nodes (N)

NX Nodes cannot be assessed

N0 No regional node metastasis

N1 Single node metastasis, 2 centimeters (cm) or less at largest point

N2 Single node metastasis, 2 cm to 5 cm at largest point, or multiple nodes, no larger than 5 cm at largest point

N3 Metastasis larger than 5 cm in any node

2.4.2.2.3 Distant Metastasis (M)

MX Metastasis cannot be assessed

M0 No distant metastasis

M1 Distant metastasis

M1a Distant lymph node(s) involved

M1b Bone(s) involved

M1c Other site(s) involved

2.4.3 Prevention

While prostate cancer cannot be prevented, measures can be taken to prevent progression of the disease. It is important for men over 40 to have an annual prostate examination. When identified and treated early, prostate cancer has a high cure rate.

2.4.4 Prostate Cancer Diagnosis

Detection of prostate cancer has been dramatically increased through improved screening programs including routine annual DRE and PSA [22].

2.4.4.1 Digital rectal examination (DRE)

DRE was explained in section 2.3.1.1.1. However, one-third of patients diagnosed with prostate cancer have a normal DRE.

2.4.4.2 PSA

Prostate-specific antigen (PSA) is a specific antigen produced by the cells of the prostate capsule (membrane covering the prostate) and periurethral glands. Patients with benign prostatic hyperplasia (BPH) or prostatitis produce larger amounts of PSA. The PSA level also is determined in part by the size and weight of the prostate

Figure 2-7 Prostate Gland Tubule shows a single normal prostate gland tubule cut in cross section. The cells lining the center manufacture prostate secretions among them PSA. There is a contour around the cells, called the basement membrane, which stops PSA and other secretions from accessing the blood stream. All of the human glands, such as the intestines, and the urinary tract are structured the same way: a secreting or absorbing layer of cells with a basement membrane to keep a tight separation between the inside and the outside. That is how the prostate makes a lot of PSA but only a tiny amount normally is found in the blood [24].

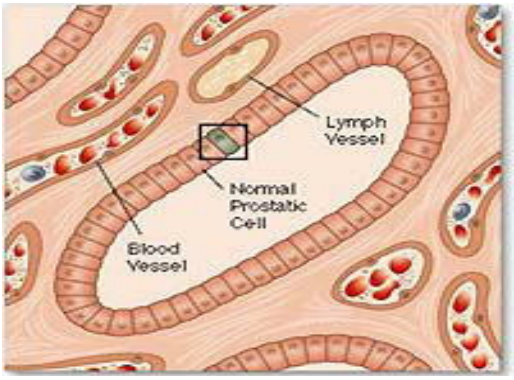


Figure 2-7 Prostate Gland Tubule

The prostate gland cells (the epithelium) are manufacturing PSA. The cells continue to manufacture PSA even if they are in locations outside the prostate. PSA levels in the blood go up if the barrier between the epithelium and the bloodstream is damaged. Three typical sources for damage are: cancer, bacterial infection, and destruction of part of the prostate by damage to its blood supply [24].

Minor elevation of the PSA levels is sometimes due to cancer, but normally a little PSA leaks from the prostate into the blood. If the prostate is enlarged then the leakage appears exaggerated. This is probably the reason for the PSA to be slightly abnormal in men with enlarged prostates who do not have cancer. The PSA level in the blood can vary by about 20% from day to day. Nevertheless, a single abnormal PSA value puts one in the higher risk for prostate cancer.

Furthermore, the use of PSA has allowed for the early diagnosis of prostate cancer as well as monitoring of patients for disease recurrence. However, there are still several clinical debates regarding the reason for the elevation in the PSA level whether it is a result of prostate cancer or a benign process [25]. Men with PSA levels more than 2.5 ng/mL have a greater than 20% chance of having prostate cancer detected by needle biopsy, while more than 50% of men with a PSA level higher than 10.0 ng/mL have prostate cancer [26,27]. However, PSA measurements have limitations

[28]. Seventy to 80% of men with "abnormal" findings on a PSA test (>4.0 ng/mL) do not have prostate cancer. Furthermore, nearly 20% of men with biopsy-proven prostate cancer have PSA values in the normal range (<4.0 ng/mL). Finally, as many as 30% of men with PSA-detected prostate cancer do not have organ-confined disease.

In a comparison between digital rectal examination and serum prostate specific antigen (PSA) in the early detection of prostate cancer the results showed that, PSA detected significantly more tumors (82%) than digital rectal examination (55%). The use of the two methods in combination increased detection of organ confined disease by 78% over digital rectal examination alone [29].

In an attempt to limit the number of benign prostate biopsies without missing a significant number of clinically important cancers Several "PSA derivatives" have been studied such as PSA velocity, PSA density, age-specific PSA and complexes PSA measurements. Currently, the most widely used is the serum percent-free PSA measurement. In general, using this measure for men with indeterminate total PSA levels (4.0-10.0 ng/mL) can save 20-30% of unnecessary biopsies and only miss 5-10% of cancers [30]. Free PSA is that percentage of the total PSA that circulates in the blood without a carrier protein. Studies suggest that malignant prostate cells produce more bound PSA; therefore, a low level of free PSA in relation to total PSA might indicate a cancerous prostate, and a high level of free PSA compared to total PSA might indicate a normal prostate, BPH, or prostatitis.

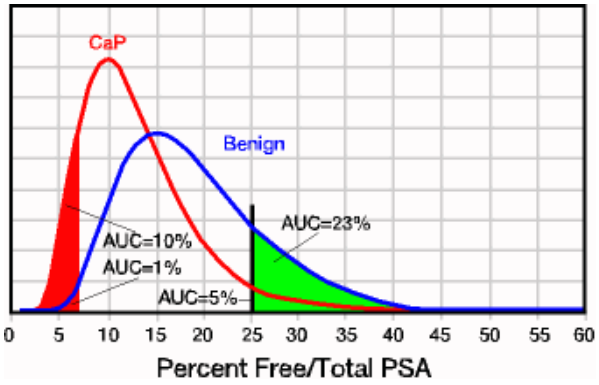


Figure 2-8 Relationship between percent Free PCA and Cancer incidences

Figure 2-8, provided by Dianon, Inc. [23], shows careful study. The horizontal axis is the percentage of free PSA. The two graphs come from analysis of 4,000 patients with abnormal total PSA between 4 and 10 ng/ml. The red line (the graph to the left), corresponding to prostate cancer patients, has a sharp peak at 10%. Most patients with prostate cancer have a free PSA less than 15%. The blue line corresponds to the distribution of free PSA in patients whose biopsies did not show prostate cancer. AUC means "area under the curve". The relative size of the area under the curves for prostate cancer and benign prostate enlargement correspond to the chances of finding prostate cancer.

From the graph, it shows that patients with free PSA below 7% usually have prostate cancer. They should undergo biopsy. If biopsy is negative, they need repeat biopsy at frequent intervals if the free PSA is reproducibly low. Patients with free PSA over 25% usually have benign prostate hyperplasia.

It is recommended that these patients undergo a single ultrasound and biopsy session. Moreover, they need their blood rechecked annually and do not repeat biopsy unless the free PSA starts to fall or the total PSA continues to rise [24]. While for the patients whose percent of free PSA values lies between 7 and 25 the decision depends mainly on the life expectancies of the patient as well as the other clinical results.

Although an abnormal PSA measurement suggests cancer, the diagnosis is always made histologically. TRUS guided needle biopsy is a safe and effective method of diagnosing prostate cancer. However, a single set of six systematic biopsies may still miss the cancer [31] and one out of five cancers will be missed in needle biopsy [32]. Some studies suggest that taking more than six biopsies at a time enhances cancer detection by as much as 30% [14, 32, 34].

It is clear from the previous discussion that only regular tests will not lead to accurate diagnosis of the Prostate cancer. As a result, physicians tend to use medical imaging techniques such as MRI (magnetic resonance imaging) and TRUS (TransRectal UltraSound) to help in cancer diagnosis. The main goal of these imaging techniques is to assist the biopsies since it helps in narrowing down the number of suspected cases of prostate cancer and recommend suspicious regions for biopsies.

Despite the improved ability of medical imaging to detect and stage malignancies (e.g., lung, colon, and breast cancer), imaging has had a minor role in the case of prostate cancer and scientific evidence indicates that many prostate cancers are undetected by medical imaging [32]. Because of a lack of appropriate technology, it has not been possible to accurately incorporate or correlate the complex disease patterns of prostate cancer into the biopsy techniques or the staging methodology, thus the accuracy of the findings estimated by the existing biopsy protocols are unsatisfactory. A major limitation of conventional biopsy protocols is the lack of a method to recommend specific locations with the highest probability of cancer occurrence for core-needle biopsies [32].

Medical imaging is another important diagnostic tool that is explained in section 2.5

2.4.4.3 Tumor Biopsy

A biopsy is performed to determine the type of cancer, its location, and stage of development. The biopsy is performed with the patient lying on his side with his knees brought up to his chest. A biopsy needle, similar to one used to draw blood, is inserted through the perineum into the tumor. A probe, guided by transrectal ultrasound (TRUS), is inserted into the rectum to help the physician properly place the needle into the tumor. The needle is projected through the tip of the probe, and a cell sample is extracted from one or several areas of the tumor into the syringe. The pathologist analyzes the sample(s) to confirm the diagnosis of a cancerous tumor and determine its type. The results are obtained within 5–10 working days. After a biopsy, blood in the urine and stool is common and usually diminishes within 1–2 weeks. Patients also experience a dull ache in the perineum for several days. Blood may appear in the semen. If the patient develops a large number of blood clots or cannot urinate, the physician should be contacted or the patient should go to the emergency room. Rarely, biopsy of a cancerous tumor may cause spreading, or "seeding," of cancer cells along the path of the biopsy needle. The biopsy operation is usually a sextant biopsy where a sample is taken from six

random regions in the gland guided by TRUS images. This operation might miss cancerous regions and lead to false negative result.

2.4.4.4 Gleason Score

DF Gleason in 1974 and 1977 published a prostate cancer grading system that is now universally applied [35]. His system grades the cancer by looking mainly at the patterns of cancerous glands.

The biopsy sample(s) is examined under a microscope for cells or groups of cells that are markedly different from healthy tissue. The greater the disparity between the healthy cells and those that are malignant, the more likely the tumor is aggressive and will spread. The pathologist examines two tissue samples taken from different areas of the tumor and assigns a score of 1 to 5 to each sample. The more abnormal the tissue is, the higher the score. The sum of the two scores produces the Gleason score. Gleason scores of 2 to 4 indicate that the tissue is not too abnormal; 5 to 7 moderately abnormal; 8 to 10 highly abnormal. Higher scores suggest aggressive tumors that likely require aggressive treatment.

Figure 2-9 BPH Cells

Figure 2-9 shows photomicrograph of benign prostate hyperplasia. It shows that the glands are large, they are uniform in size and they are separated from each other by tissue called stroma. However, Prostate cancer, as seen under the microscope, consists of glands, which are too small, too close to each other. The individual cells are also abnormal, with enlarged nuclei.

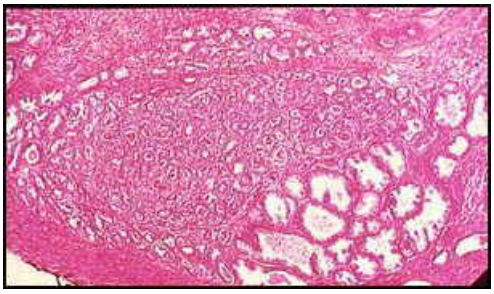


Figure 2-10 Gleason Score (1+1)

The lowest grade tumor is Gleason 1+1 as shown in the Figure 2-10 where a nodule of Gleason 1 prostate cancer, with a pattern of small uniform glands organized into a nodular pattern is shown.

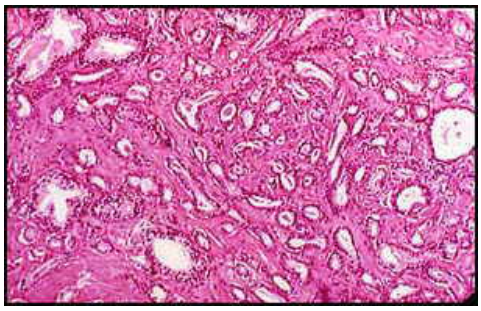


Figure 2-11 Gleason Score = 3

Figure 2-11 shows Gleason grade of 3. The glands are irregularly shaped. They are mixed in with some normal glands.

2.5 Medical Imaging

Traditionally, medical imaging is mainly categorized based on two criteria; the first depends on the image representation while the second is based on the application.

2.5.1 Representation-based

Three different image representations are the most common:

- 2-D: Is the typical image representation. The image is represented as a function of location in two dimensions, with its intensity mapped as a gray-scale.
- 3-D: This image is constructed of a set of 2-D images as a function of location within the body (a stack of slices).
- 3-D + time: typically, a set of 2-D images (a stack of slices) taken repeatedly over time.

2.5.2 Purpose-based

Purpose based imaging can be divided into three different groups based on the type of application

2.5.2.1 Screening (detection)

The goal in screening is to detect conditions that may be disease or evidence of disease, with the intent of conducting a detailed follow-up of suspicious findings. Consequently, the emphasis is usually on sensitivity (the detection of abnormalities when they exist) at the cost of an increased false-positive rate (decreased specificity). Given that, screening is performed on large number of people, it is important that the procedures be low in cost and provides quick results. The features

should, therefore, be easy to extract, contribute to the sensitivity of the procedure, and require minimal user intervention. Examples of images used in screening include x-ray mammograms for breast cancer [36], structured visible light images for childhood scoliosis (curvature of the spine) [37], and photography for diseases of the retina of the eye [38].

2.5.2.2 Diagnosis (classification)

Diagnosis aims to make a specific identification of a problem: for example classify the suspicious region in the breast is a benign breast tumor, or a carcinoma [39]. Additional features may be required in cases where screening preceded diagnosis [40]. Another important application is comparing two images representing the same region, but taken at different instants (to describe changes in a condition), or comparing the given image to an atlas or other standard, for diagnosis or reporting [41].

2.5.2.3 Therapy and Treatment Planning

In radiation oncology, it is often necessary to identify and align structures in two images: one is the "prescription image" used to indicate the areas to be treated [42]; the other is the "portal image," taken just prior to treatment. The portal image is intended to confirm that the treatment beam is aimed correctly [43]. Typically, however, image quality is low, therefore, various enhancement methods [44] have been employed, but it is nevertheless often necessary to extract features (e.g., shapes, areas) to provide the basis for identification of the treatment areas and their boundaries [45].

2.6 Medical Imaging modalities used for Prostate Cancer diagnosis

2.6.1 MR Imaging

Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopic Imaging (MRSI) are recently introduced as additional modalities for the local evaluation of prostate cancer [46] in an attempt to overcome the limitations of traditional prostate cancer diagnosis methods,

MR imaging has not been considered a primary screening method for detecting prostate cancer, although its use in men with previously negative findings at biopsy is being studied. The main goal of MR imaging for prostate cancer is to detect extracapsular spread of tumor. When good-quality endorectal coil MR imaging is performed, it is highly sensitive (80-90%) for the detection of extraprostatic disease [47]. MR imaging has many attractive features. For example, MR spectroscopy is considered one of the most promising areas of prostate imaging research [48]. MR spectroscopy has been shown to improve detection, localization, and staging of prostate cancer as well as to detect recurrent disease after therapy [49].

Several other MR imaging techniques are also promising. These include dynamic contrast-enhanced MR Imaging in which tumor angiogenesis can be assessed [50], and T2*-weighted imaging. Tissue elastography is another potential functional tool provided by MR imaging [51]. These techniques may provide a more comprehensive evaluation of tumor aggressiveness than anatomic MR imaging now

provides. However, Compared with TRUS, endorectal MRI with a 0.5 T magnet provides greater sensitivity and specificity for capsular penetration and increases sensitivity for seminal vesicle invasion [52]. A limitation of MR spectroscopy is it's technical demands, requiring specialized software and expertise in obtaining and interpreting spectral peaks. Currently the main clinical indication of MRI is nodal and bone metastasis [53].

2.6.2 Nuclear Medicine

Nuclear medicine such as bone scans provide poor spatial resolution but are highly sensitive to metabolic activity. The most recent dramatic forays into prostate cancer imaging have been with indium-111 capromab pendetide, a prostate specific membrane based antibody labeled to indium [54] whose use is still very limited [55].

Fluorine-18-fluorodeoxyglucose positron emission tomography is having a new prospective in assessing metabolism to detect cancer. The specific uptake value is a reproducible index of metabolic activity and reliably that predicts the presence of cancer, at least in cancer of the lung and colon. However, fluorodeoxyglucose positron emission tomography has proven to be disappointing in prostate cancer [56]. A bone scan is used to detect metastesis to bones. It is not used in patients with small cancers and low PSA levels.

2.6.3 CT Scans and PET Imaging

Computer-assisted tomography (CAT scan or CT scan) is an x-ray procedure that produces three-dimensional images of internal organs and glands. It can be used to detect pelvic lymph nodes enlarged by cancer, but results may be insufficient for diagnosis. CT scans are used only when tumors are large or PSA levels are high. However, the CT is used mainly for the radiation oncology procedure [57, 58]. It is shown in [57] that daily CT localization is a precise method to improve daily target localization in prostate cancer. However, it requires significant human and technical resources that limit its widespread applicability. Conversely, localization with the ultrasound system is simple and able to image the prostate at the treatment machine in the treatment position. This study revealed ultrasound targeting to be functionally equivalent to CT. in another study that is used to compare between the CT scan and the MRI in defining the radiotherapy treatment, it was shown that the MRI obtained better segmentation results than the CT scans. This study showed promising results for the use of MRI for treatment planning [58].

Another imaging modality is also introduced for imaging the prostate that is used in the oncology application, which is the Positron Emitted Tomography (PET); it showed promising results but is not applied for diagnosis yet [59]. Because of the higher cost, CT and MRI are to be used only for metastasis detection as recommended in [53]

2.6.4 TRUS

TRUS is typically used to measure the size of the prostate and visually identify tumors. A probe inserted into the rectum emits ultrasonic impulses against the prostate. The images are projected on a monitor, so the physician can examine the gland and surrounding tissue for tumors. Since TRUS is able to accurately reveal the peripheral zone—the region where most cancers develop—it is advocated for prostate cancer screening. However, it became clear that sonography lacked both the sensitivity and the specificity to be used for early detection. Currently, sonography is primarily used for screening and to help guide needle placement for transrectal systematic prostate biopsies. However TRUS is favored than both MRI and CT scans because of its simple and small instrumentation where the machine can be transferred to the patients' room. Moreover, it doesn't need special setup in the hospitals and it allows for real time image visualization.

2.6.5 Doppler Ultrasound

It is well known that to allow growth and infiltration, the tumour needs sufficient oxygen and nutrients to be supplied by the surrounding blood vessels. Therefore, it is assumed that the tumour changes the metabolism and the environment of its surrounding. This change in local blood vessels can be studied by Doppler ultrasound [53]. Color Doppler sonography depicts flow within the prostate gland and can better locate tumors that are otherwise invisible on gray-scale imaging. However, color Doppler signal alone often lacks sufficient specificity [60]. Doppler US has improved positive prediction rate, however, there are some limitations for Doppler imaging for cancer detection. Since prostate cancer is not highly vascular, then Doppler ultrasound is insufficient for determining a biopsy strategy [61]. It is also shown in [53] and [61] that the Doppler US is poor in its sensitivity and it should be used in conjunction with TRUS.

2.7 Elastography

A new prostate cancer diagnostic modality is elastography or strain imaging that visualizes the local stiffness of tissues using TRUS. It was first introduced in 1991 and introduced to prostate cancer diagnosis in 1998 [62, 63], however it still lacks real time capabilities. A fast algorithm was introduced in 2000 [64, 65] that allows elastography to be used in adjunct to DRE, PSA and TRUS for prostate cancer diagnosis in certain clinics. Yet elastography is still under clinical evaluation and not used widely.

2.8 Computer Aided Diagnosis (CAD) using TRUS

Transrectal Ultrasound (TRUS), introduced in 1971 [35], provides information about the size and shape of the prostate. In the late 1970's and early 1980's the technology progressed, allowing two clearly distinct zones to be identified within the gland, but was deficient in detecting tumors. In the mid 1980's, higher frequency transducers were introduced, resulting in higher image resolution and better display of zonal anatomy. The use of ultrasound is motivated by its proven clinical ability, its

safety record, its real time visualization and its simplicity to use. Therefore, TRUS has become the dominant imaging modality for diagnosis of prostatism, detection and staging of prostate cancer.

2.8.1 Ultrasonic Imaging

Ultrasonic systems use transducers that act as transmitters and receivers in a pulse echo mode. A short voltage pulse excites the transducer and it generates a brief ultrasound pulse. Each pulse comprises a series of alternating compressions that propagate through the body. As the pulse travels, its pressure amplitude is progressively decreased by absorption and is partially scattered by tissue density changes and propagation velocity. Tissues usually produce weak scattering so that most of the ultrasonic energy continues traveling deeper in the body. A part of the backscattered pressure pulses propagate back to the transducer, which integrates those echo pulses, and generate the RF signal. The envelope of the RF signals is used to create the images along a line corresponding to the transducer orientation [66]. The image resolution increases as the frequency increases, however tissue attenuation increases with the frequency. Accordingly, deep organs that require deep penetration are imaged using low frequencies, while sections of the eye are examined at high frequencies to achieve high resolution.

2.8.2 Modes of TRUS data analysis

Although ultrasound imaging is a popular tool for prostate cancer diagnosis, it results in poor accuracy in early prostate cancer diagnosis. In addition, inaccurate ultrasound images might lead to lots of misled biopsy operations to be performed, and to a random needle biopsy locations which might miss the cancerous regions. CAD is being investigated in order to obtain the best out of the TRUS technology. CAD using ultrasound imaging is based on extracting parameters from the returned echoes in order to identify the tissue type present in the image. The first parameter is RF signal that is extracted from the returned echoes prior to machine processing. This parameter was used for prostate tissue characterization using spectral features where the power spectrum of the RF signal is linearized and some features such as y-intercept and mid-band are extracted from the linearized power spectrum [66, 67, 68]. The second parameter is the image texture where features are extracted from the image after the echo processing is performed. Such texture parameters include the statistical characteristics of the grey level distribution in the image. This parameter was applied to prostate cancer diagnosis [69, 70, 71].

Using the RF signal has the advantage of being free from the machine processing distortion. While for the image texture applications, data is easier to collect and implement visually.

2.8.2.1 CAD using RF signals

The B-mode Ultrasound images might suffer from some degradation due to the coherent nature of ultrasound that causes speckle noise. Therefore, the RF signals are used when available for some diagnostic applications. RF signals that are acquired at the transducer terminals are used to avoid the non-linear transformation effects from the RF signal to the image form. Spectral techniques have been mainly deployed in two modes; the first calculates the average spectral parameters within a

demarcated spatial region. This mode is mainly used in database studies that elucidate parameter values indicative of specific disease [66, 67]. The other mode generates spectral parameter images that can be used for pattern recognition where a separate image is generated for each parameter (midband image, slope image) [66]. These images are used to differentiate between different tissues in the organ.

2.8.2.2 CAD using TRUS Images

Computer Aided Diagnosis (CAD) for prostate cancer requires four major steps: segmentation, ROI identification, feature analysis, and classification. The accurate detection of prostate boundaries from ultrasound images (Segmentation) plays an important role in several applications such as the measurement of prostate gland volume. Many efforts have been done in the segmentation process, which makes it well established [72], [73]. ROI identification is highlighting the most probable cancerous regions in the gland, a step that is normally achieved with the help of expert radiologist. This step is crucial as studying the whole image will lead to distorted features and will not reflect the medical condition. Promising results are obtained using Gabor multi-resolution analysis in [71].

Feature analysis is mainly constructing features from the TRUS image as well as selecting the most discriminatory features among the constructed ones. The constructed features could be statistical features, spectral features or model based features. Co-occurrence matrices are the earliest used feature construction methods in [69] for prostate cancer diagnosis. Where the second order features using co-occurrence matrices are constructed and analyzed to rank these features according to their discriminatory power. While in [70] a combination of statistical features as well as features constructed from the RF data are used for prostate cancer diagnosis. In some cases, the constructed features might have some redundant and correlated information therefore feature selection algorithms are applied for the first time for prostate tissue characterization using TRUS images in [71]

Following that, is the classification stage, which depends mainly on the quality of the constructed / selected features and the precision of the classifier. Tissue typing in the TRUS image attempts to differentiate the cancerous and the non-cancerous regions. Different classifiers have been applied for this task such as ANN in [15], Fuzzy Inference System (FIS) in [70], KNN, DT and SVMs in [71].

2.9 Conclusions

In this Chapter an over view of the prostate gland anatomy, its different diseases as well as the different diagnosis methods were given. Further, different cancer diagnosis and staging schemes were summarized and different imaging modalities used for prostate tissue characterization were briefed. Furthermore, CAD using TRUS data such as RF signals or TRUS images were discussed.

Chapter 3
ROI Identification

3.1 Preamble

Given the outlined TRUS image (only the gland) for the suspected patient, the first step in the proposed CAD system is to identify the suspicious ROIs. This task is achieved using multi-resolution analysis by applying Gabor filters that are capable of segmenting textured images according to the frequency response of its pixels. The typical texture segmentation assigns to each pixel a feature vector representing the response of that pixel to each of the filters in the Gabor filter bank. A simple clustering algorithm such as k-means is then used to segment the image according to each pixel's feature vector. The pixels that have similar response to the same filters are assigned to the same cluster. This texture segmentation process was successfully applied to segment some test images to several regions that are then compared to the gold standards. A modified version of the texture segmentation algorithm is designed in this chapter and then applied to the TRUS images where the suspected ROIs are segmented and highlighted. At this stage, the doctor expertise is incorporated in the decision making to identify those regions. The expertise information is achieved by examining the radiologist identified regions and extracting important information from these regions such as the length of the contour of the selected regions as well as their frequency level. This information is then integrated to the ROI segmentation algorithm, which reduces the number of candidate regions that are then considered for further analysis. The process of using multi-resolution analysis to identify the ROIs is explained in the following sections.

3.2 Typical Gabor Texture Segmentation Process

Multi-channel filtering is an excellent method for texture investigation. By processing the image using multiple resolution techniques, it is decomposed into appropriate texture features that can be used to classify the textures accordingly [74].

From the fact that texture exhibits a sort of periodicity of basic patterns, it can be concluded that different objects are identified based on their texture properties. The Human Visual System (HVS) is formed of a number of independent narrow band filters tuned to different frequencies. In the HVS system, people use three features for texture analysis, mainly repetition, directionality and complexity. Repetition and directionality are represented by frequency and orientation, while complexity relates to the complexity of texture. Therefore, a model for the HVS interpretation of texture has been developed based on multi channel filtering of narrow bands. Simple cells in the visual cortex were found to be sensitive to different channels of various spatial frequencies and orientations. Since texture repetition can be represented by frequency and directionality by orientation, the multi-channel filters are expected to mimic the behavior of the human visual system (HVS) [74].

Gabor filters have been successfully applied for many different applications such as segmentation of textured images [74, 77], edge detection [78] and retina identification [79]. However, the design of these filters remains an open issue [80]. This is because there are several approaches used to design the filter bank and lots of debate is present between these approaches. One of these approaches employs a large bank of filters [74] where the filter parameters are predetermined by an ad-hoc procedure and are not necessarily optimized for a specific task. Yet, some segmentation tasks may not require all these filters for better performance or may not bear the computational burden imposed by the large filter bank. In addition, the filter bank produces a feature vector that is usually high in dimension. In contrast, the second filter design approach is based on designing and optimizing the filters to fit a pre-existing segmentation problem [80].

The typical process of texture segmentation using Gabor filters is based on the first approach. When generating texture features using multi-channel filters, two stages should receive special consideration. First, the characterization of the filters (e.g. parameters, number, orientation, and spacing) must be selected carefully. Second, feature extraction from the raw filter outputs (filtered images) should be performed to improve the feature set.

Therefore, Gabor filter based segmentation process typically consists of the following steps:

- It first involves a proper filter bank design that should be tuned to different spatial frequencies and orientations to cover the spatial frequency space.
- The image is then decomposed into a number of filtered images corresponding to different filters,
- Next, the features are extracted from these images.
- Finally, the pixels are clustered in the feature space to produce segmented image. Figure 3-1 shows a schematic diagram of the texture segmentation process [75, 76].

3.2.1 Gabor Filter Implementation

When generating texture features using multi-channel filters, two primary issues must be addressed. The first issue deals with the functional characterization of the channels as well as their number, orientation and spacing. The second issue deals with extracting significant features by data integration from different channels.

Gabor filters have the ability to perform multi-resolution decomposition due to its localization both in spatial and spatial-frequency domains. Texture segmentation requires simultaneous measurements in both the spatial and the spatial-frequency domains. Filters with smaller bandwidth in the spatial-frequency domain are more desirable as they allow finer distinctions [74]. On the other hand, accurate localization of texture boundaries requires filters that are localized in the spatial domain. However, normally the effective width of a filter in the spatial domain and its bandwidth in the spatial-frequency domain are inversely proportional. The Gabor filter is well suited for this kind of problem as it optimizes the bandwidth in both domains [78]. The use of even symmetric Gabor filters was

recommended in [81] for complex object detection, a problem that is close to TRUS image segmentation.

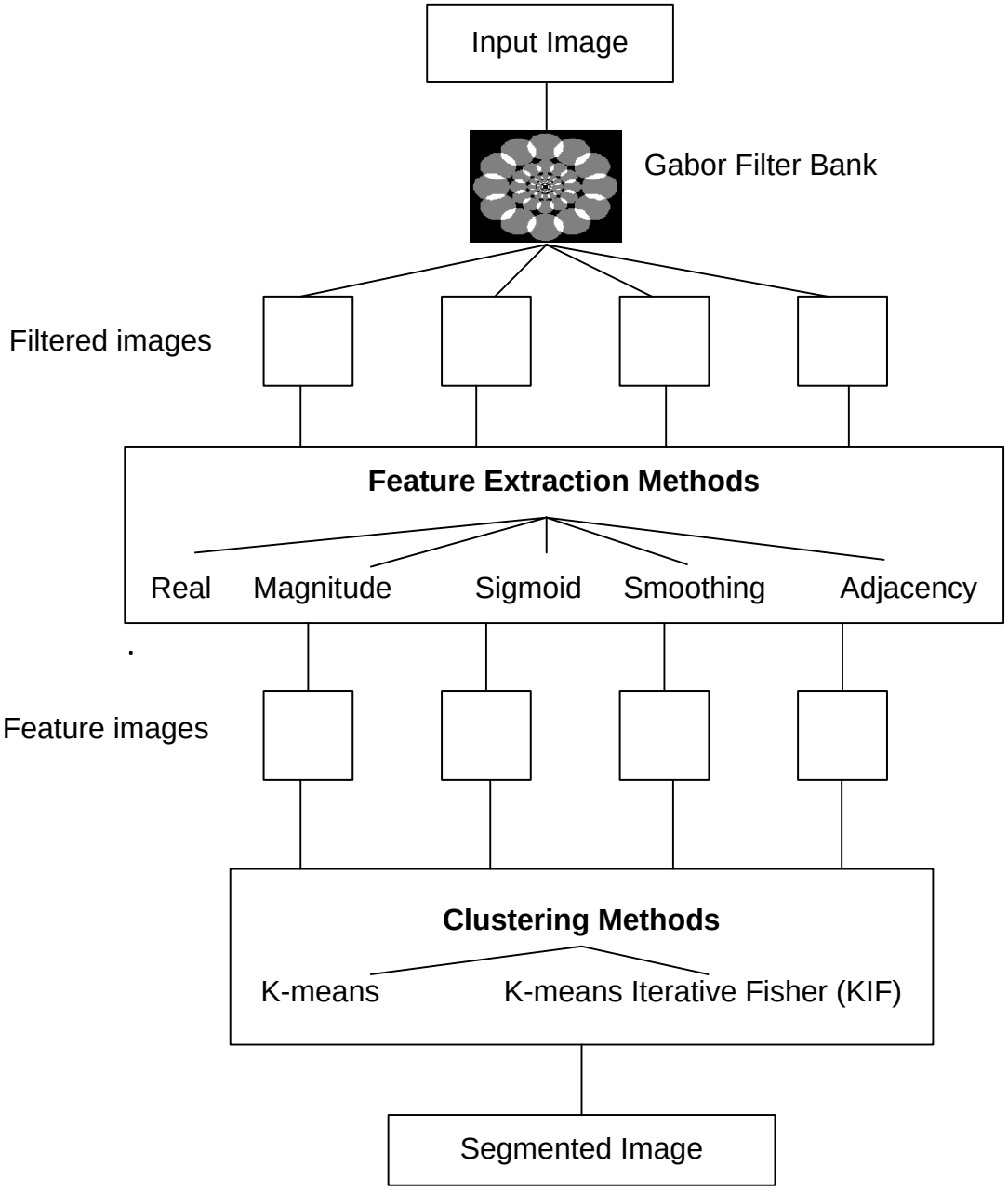


Figure 3-1 Image Segmentation Using Gabor Filter Bank

In terms of functionality, the Gabor function in the spatial domain is a Gaussian modulated sinusoid. For a 2-D Gaussian curve with a spread of σ_x and σ_y in the x and y directions, respectively, and a modulating frequency u_0, the real impulse response of the filter is defined in [74, 76] and is given by:

$$h(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^{\,2}}+\frac{y^2}{\sigma_y^{\,2}}\right]\right\}\cdot\cos(2\pi u_0 x+\varphi) \qquad \textbf{3-1}$$

In the spatial-frequency domain, the Gabor filter becomes two shifted Gaussians at the location of the modulating frequency. The equation of the 2-D frequency response of the filter is given by:

$$H(u,v) = \exp\left\{-2\pi^2\left[\sigma_x^2(u-u_0)+\sigma_y^2v^2\right]\right\}+\exp\left\{-2\pi^2\left[\sigma_x^2(u+u_0)+\sigma_y^2v^2\right]\right\} \qquad \textbf{3-2}$$

The selection of σ_x, σ_y determines the resolution in both the spatial domain and the spatial frequency domain. Low values of σ_x, σ_y lead to better spatial resolution while their high values favour frequency resolution. Usually on segmenting an image, short spatial intervals are favourable in order to be able to approximate the boundary between different textures. However, smaller frequency bandwidths are preferable in order to facilitate texture discrimination. Unfortunately, the spatial extent and the spatial frequency bandwidths are inversely proportional. These conflicting goals of simultaneous spatial and frequency resolutions are referred to as the *Uncertainty* principle and are explained in [75]. A Gabor function has an interesting property, which is the optimal joint resolution in both the spatial and the spatial frequency domains; this property is explained in [82] and is summarized as follows:

If Δx represents the resolution in the spatial domain and Δu represents the resolution in the spatial frequency domain then the joint resolution is bound by the inequality:

$$(\Delta x)(\Delta u) \geq \frac{1}{4\pi} \qquad \textbf{3-3}$$

The one dimensional Gabor function achieves the equality of this formula, the case for 2-d Gabor function is shown in [82].

Moreover, the position (u_0, φ) of Gabor filters in the frequency domain is supposed to be carefully setup to capture essential texture discrimination information. This is an important issue because centre frequencies of channel filters must agree very closely with the emergent texture frequencies. Otherwise, the channel frequency responses fall off too rapidly within a small neighborhood of the centre frequencies. If achieving this task is not possible; then the global solution is to spread filters throughout the frequency domain field to capture salient information. By providing near uniform coverage of the spatial-frequency domain with Gabor filters, the problem of selecting central frequencies is avoided. Jain and Farrokhnia [75] used a bank of even symmetric Gabor filters to characterize the channels. This is represented in equation 3.2 where two Gaussians appear in the spatial-frequency domain. The frequency u_0 and the rotation angle φ define the centre location of the filter. By tuning u_0 and φ to different center locations, multiple filters that cover the spatial-frequency

domain are created. The output images from the filters are then used for texture segmentation that depends mainly on the response of the pixels to each of the filters.

3.2.2 Texture Segmentation

The process of texture segmentation used for the test images is summarized in the following steps, which will be explained in the following sub-sections:

- Filter bank design.
- Decomposition of the input image using the filter bank.
- Feature extraction.
- Pixels clustering in the feature space.

3.2.3 Filter Bank Design and Decomposition of the Image

The parameters of the filter should be chosen by selecting a set of frequencies and orientations that cover the entire spatial-frequency space, and capture texture information as much as possible. Two orientations are tested in this work for the Brodatz images: the first is 30^o as recommended in [74], and the second is 45^o as recommended in [75]. It is expected that the 30^o will give better results as we do finer orientation, which means capturing more features than the 45^o orientation. However, the candidate method is more expensive. The frequencies that are used for filters are inspired from [74]:

$$1\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, ...(N_c/4)\sqrt{2} \quad \text{Cycles /image width} \quad \text{(CPP)} \qquad \textbf{3-4}$$

For the case of 30^o, the number of filters required is more than that in case of 45^o for the same image width. The filters with very low frequencies ($1\sqrt{2}, 2\sqrt{2}$) can often be left out as they capture spatial variations that are too large to be considered texture [75]. The variable σ_x is determined by setting B_F (frequency bandwidth) to 1 as recommended in [74].

$$\sigma_x = \sigma_y = \frac{\sqrt{\ln 2}(2^{B_F}+1)}{\sqrt{2}\pi F(2^{B_F}-1)} \qquad \textbf{3-5}$$

The above-proposed parameters are used in this work so that the filter can accurately capture the texture information. Centre frequencies of channel filters must be close to the characteristic texture frequencies, otherwise the filter responses will fall off rapidly [74]. A representation of the filter bank that is used in this work results in nearly coverage of the frequency domain as shown in Figure 3-2.

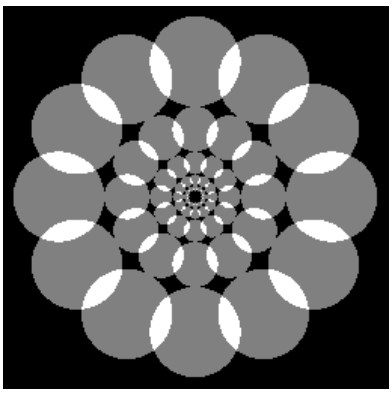


Figure 3-2 The filter bank in the spatial frequency domain

By applying this filter bank to the image, it results in decomposing the image to a number of images that are equal to the number of filters used. Each image corresponds to the response of the pixels to one of the filters used. These images are composed of a real part and an imaginary part. Feature extraction from these filtered images is important in order to get the information needed for segmentation. Different feature extraction methods are explained in the next section.

3.2.4 Image Feature Extraction

Filter outputs by default are not appropriate for identifying key texture features. A number of feature extraction methods are proposed in literature to extract useful information from the filter outputs. Some feature extraction methods are reviewed in [74] and some of which are tested on the test images in this work that mainly include:

1. Using the ***Magnitude Response***, where the texture identification can be performed based on the magnitude of the output of the Gabor functions. In the case of a filter that matches the particular texture, the magnitude of the output is large to allow identification.

2. Applying the ***Spatial Smoothing*** where Gaussian smoothing is known to improve the performance of Gabor filters for texture analysis. Post filtering the channel amplitudes with Gaussian filters having the same shape (standard deviation) as the corresponding channel filters is recommended in [76].

3. Using only the ***Real Component***: Jain and Farrohknia [75] used the real component of the images obtained by a bank of even symmetric Gabor filters to characterize the channels.

4. Using ***Pixel Adjacency Information*** is suggested in [75] where using this method is valid for synthesized images due to the fact that pixels belonging to the same texture are close to each other, so they should be clustered together. However, this will not perform well in the case of TRUS images as the cancerous region might be small, therefore, applying this method will blend the cancerous region into the neighboring regions.

5. ***Sigmoidal Function***: Using a Non Linear Sigmoidal function that saturates the output of the filter where each filtered image was subjected to a Sigmoidal non-linear transformation [75] that can be interpreted as a blob detector. It is indicated by:

$$\varphi(t) = \tanh(\alpha t) = \frac{1 - e^{-2\alpha t}}{1 + e^{-2\alpha t}} \qquad \textbf{3-6}$$

Where α is an empirical constant, α = 0.25. The logic presented in [75] was that most textures can be characterized by blobs of different sizes and orientations.

At the end of the feature extraction step, a set of feature images extracted from the filter outputs is available for clustering. Pixels that belong to the same texture region have the same texture characteristics and should be close to each other in the feature space (their response to the Gabor filters will be close to each other). The final step for the texture segmentation is unsupervised clustering where pixels are clustered into a number of clusters representing the original texture regions. Labeling each cluster will yield to segmented image. K-means clustering algorithm is used for that purpose.

3.2.5 Clustering in the Feature Space

At the end of the previous stage, we are usually left with a set of feature images extracted from the filter outputs. Pixels that belong to the same texture region have the same texture characteristics, and should be close to each other in the feature space (i.e., their response to the Gabor filters will be close to each other). The final step in unsupervised texture segmentation is to cluster the pixels into a number of clusters representing the original texture regions. Labeling each cluster usually yields to the segmented image.

Different approaches were applied earlier for the image clustering process such as K- means [81] and K-means iterative Fisher (KIF) [80]. The k-means clustering algorithm is used for the test Brodatz images in this work for simplicity. However, this means we have to provide the algorithm with the number of clusters beforehand, which means the number of different textures in the image is previously known.

K-means starts by assigning the cluster centers to random points in the input set. Then it starts calculating the distance from each point to its nearest cluster center- based on the Euclidean distance- the next step is to recalculate the cluster centers as the mean of each cluster. The algorithm works iteratively by assigning the points to their nearest cluster center, updating the cluster centers until it converges and no more changes can be made. When clustering is done, each pixel is labeled with its respective cluster, finally producing the segmented image.

3.3 Experimental Results

Gabor Filter outputs by default are not appropriate for identifying key texture features. A number of feature extraction methods are proposed in literature to extract useful information from the filter outputs, some of which include:

1. Using the Magnitude Response,
2. Applying the Spatial Smoothing
3. Using only the Real Component of the output image
4. Using Pixel Adjacency Information feature due to the fact that pixels belonging to the same texture are close to each other, so they should be clustered together. However, this will not perform well if there are some texture regions that are not adjacent in the image.
5. Using the non-linear Sigmoidal function.

The effects of these five methods are examined in this section. Where spatial smoothing can be applied to any of the above-mentioned methods, and is well known to enhance the performance of the segmentation processes as it suppresses large variations in the feature map in areas, which belong to the same texture. However, too much smoothing can have a negative effect on the localization of texture region edges. Each filter output is smoothed using a Gaussian smoothing function that matches the corresponding filter spatial Gaussian curve [80].

Figure 3-3 shows a five texture image taken from the Brodatz texture album [83]. Figure 3-4 and Figure 3-5 show a comparison between the different feature extraction methods applied to this image.

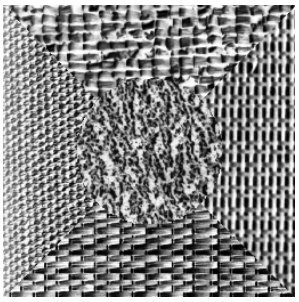


Figure 3-3 Five texture Brodatz image

From Figure 3-4 and Figure 3-5, it can be subjectively argued that the smoothed magnitude response will perform better than the other methods as it captures most of the right side texture information while, on the other hand, the real and smoothed real response as well as the Sigmoidal function doesn't differentiate between different image textures.

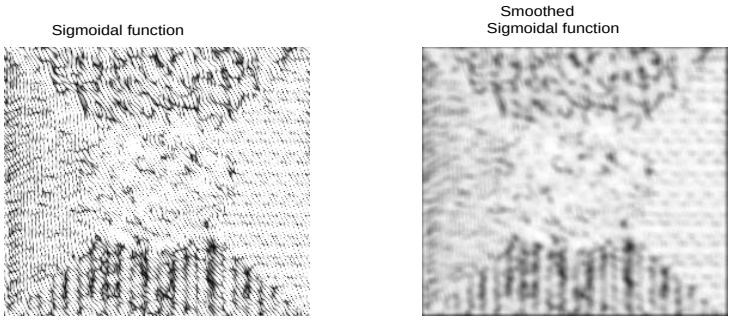


Figure 3-4 Applying the Sigmoidal function to the filter output

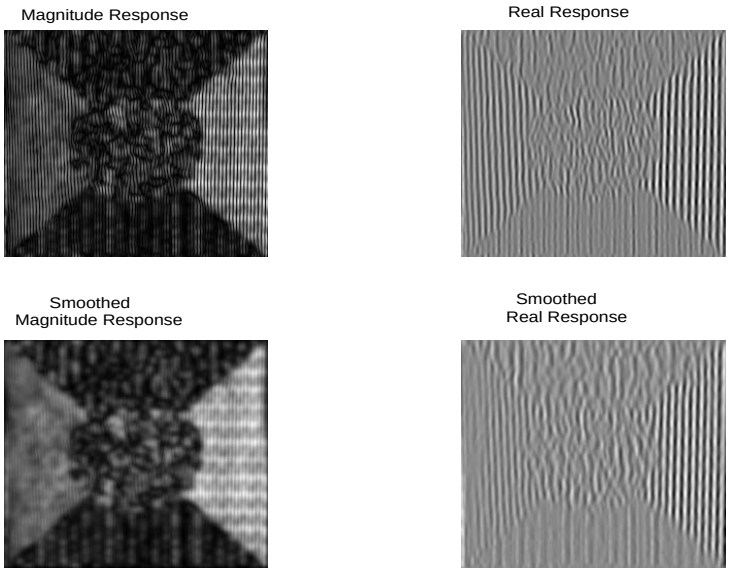


Figure 3-5 Using the magnitude and the real response from the filter output

3.3.1 Brodatz Images Segmentation Results

The multi-channel approach discussed in the previous sections is implemented and tested against a number of texture images from the Brodatz album due to their popularity in this area. Figure 3-6 shows the segmentation result of the Nat-5c five-texture image from the Brodatz album. The figure shows the segmentation based on orientation separation of filters of 30^{o}. Figure 3-7 shows the segmentation based on orientation separation of filters of 45^{o}.

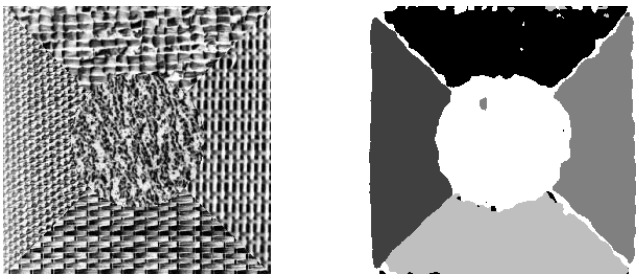


Figure 3-6 Segmentation of Nat-5c image using magnitude response and 30^o orientation separation.

It is clear from the results that the filter set at 30^o separation is performing much better than the other set. This was expected as at the 30^o orientations we can cover more of the spatial frequency space, which results in capturing more texture information. It is also observed that the texture boundaries are well localized to some extent. However, we do not get sharp localization due to the smoothing that was done as a post-processing step to the magnitude of filter outputs. To visualize the effect of smoothing response, segmentation of the same image without smoothing was tested and the result is shown in Figure 3-8.

Figure 3-7 Segmentation of the Nat-5c image using the magnitude response for feature extraction using 45^o orientation separation

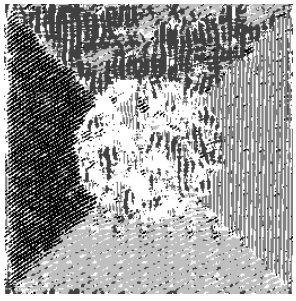


Figure 3-8 Effect of non-smoothed segmentation.

The effect is clear and shows that smoothing suppresses the variations in the texture features within the same texture. Non-smoothed segmentation is severely affected by this variation and the result suffers non-contiguous labeled regions opposing to the case of the smoothed segmentation.

Segmentation based on different feature extraction methods is also tested in this work on four texture image. The methods used are:

- Magnitude response.
- Real response.
- Pixel adjacency information.

In addition, the effect of the orientation separation is also tested in this work. The results of this test are shown in Figure 3-9, Figure 3-10, Figure 3-11 and Figure 3-12. The magnitude response gives the best result when combined with pixel adjacency information, as excellent segmentation of the four texture image was achieved.

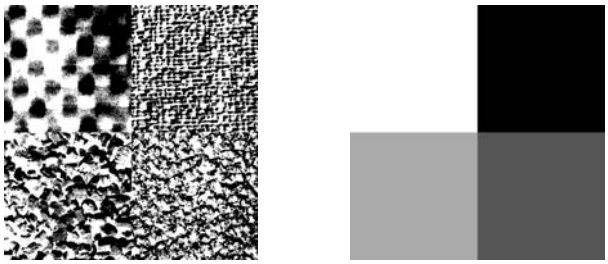


Figure 3-9 The original texture image and the segmented one using Magnitude response, orientation of 30, and pixel adjacency.

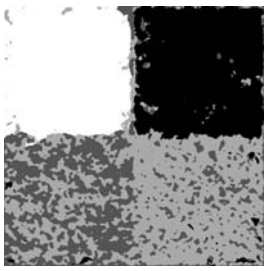


Figure 3-10 Segmented image using magnitude response, orientation of 300 , and no pixel adjacency.

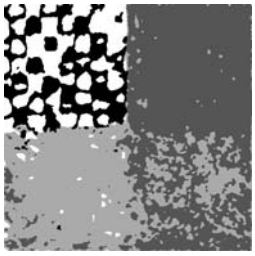


Figure 3-11 Segmented image using magnitude responses, orientation of 45, and no pixel adjacency.

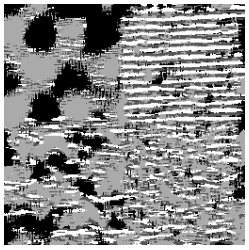


Figure 3-12 Segmented image using real response, orientation of 30, and no pixel adjacency

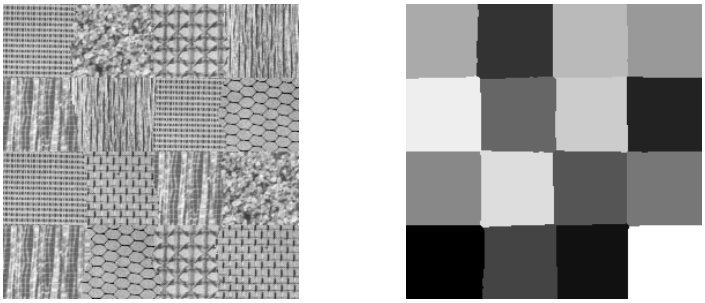


Figure 3-13 16 texture image and segmented image using Magnitude response pixel adjacency and 30^o orientation.

For the previous figures, it is shown that the magnitude response produces better results especially when using an orientation of 30^o. When it is combined with pixel adjacency, it gives excellent segmentation for the four and sixteen texture images, which is clear from Figure 3-13 . In the case of the five texture image, the implemented algorithm for the pixel adjacency is not suitable as it consideres only the vertical and horizontal neighbors to affect the pixel. However, this image still gives an excellent segmentation result using the magnitude response and orientation of 30^o, which is clear from Figure 3-14.

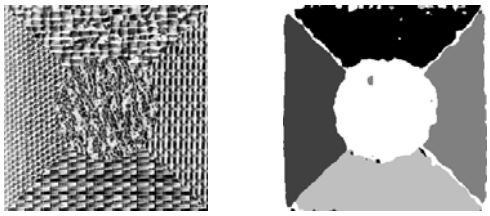


Figure 3-14 Segmentation of Nat-5 image; smoothed magnitude response and orientation of 30^0

3.4 TRUS ROI Segmentation

The results from the above section were very promising. The same technique was applied for the TRUS images with a slight change. For the TRUS images, it was found that there is no need for filling the frequency domain with filters as the changes in the TRUS images usually occur at high frequency. Therefore, the proposed algorithm offers only one Gabor filter located at high frequency for this specific application. To save the computation effort in order for the algorithm to be applied online during performing the imaging process, only the filtered image resulting from that filter is then segmented.

3.5 Gabor Filter Design for TRUS Images ROI Segmentation

The Gabor filter proved providing simultaneous optimal resolution in both the spatial domain and the spatial frequency domain. The typical symmetrical even Gabor filter was explained in detail in section 3.2.1. In this work, the same design scheme for the Gabor filter was applied successfully for the Brodatz test images. Therefore, in the proposed algorithm the even symmetric Gabor filter was tailored to fit the TRUS image segmentation problem. The impulse response of the even symmetric Gabor filter oriented along the x-axis is given by:

$$h(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{1}{2}\left[\frac{x^2}{{\sigma_x}^2}+\frac{y^2}{{\sigma_y}^2}\right]\right\}\cdot\cos(2\pi Fx) \qquad \textbf{3-7}$$

It is represented in the spatial frequency domain by:

$$H(u,v)=\exp\left\{-2\pi^2\left[\sigma_x^2(u-F)+\sigma_y^2v^2\right]\right\}+\exp\left\{-2\pi^2\left[\sigma_x^2(u+F)+\sigma_y^2v^2\right]\right\} \qquad \textbf{3-8}$$

The filter results are sensitive to selecting the filter parameters that are set using an ad-hoc method that optimizes the matching between the radiologist marked regions and the Gabor filter segmented regions. These filter parameters are:

- ***Frequency and Angular bandwidths:*** Considering the goal of our work to mimic the radiologist's opinion and augment his decision, there is a great need to try to follow the

HVS. Therefore, the frequency and angular bandwidths B_F and B_θ are set to constant values to match the psycho-visual data.

- ***The filter frequency*** *F* is set to the biggest possible value shown in [75] due to the fact that the TRUS images practice relatively high frequency changes.

$$F \cong (N_c/4)\sqrt{2} \quad \text{Cycles/image width} \qquad \textbf{3-9}$$

- ***The variable*** σ_y is set equal to the variable σ_x in order to obtain a circular Gaussian to obtain a constant spatial extent in all directions. These variables are calculated by setting B_F to 1 octave.

$$\sigma_x = \sigma_y = \frac{\sqrt{\ln 2}(2^{B_F}+1)}{\sqrt{2}\pi F(2^{B_F}-1)} \qquad \textbf{3-10}$$

$$\text{Therefore, } \sigma_x = \sigma_y = \frac{3\sqrt{\ln 2}}{\sqrt{2}\pi F} \qquad \textbf{3-11}$$

A sample Gabor filter was implemented for visualization purposes using the above parameters. It is shown in the spatial-frequency domain and the spatial domain is shown in Figure 3-15 and Figure 3-16 respectively.

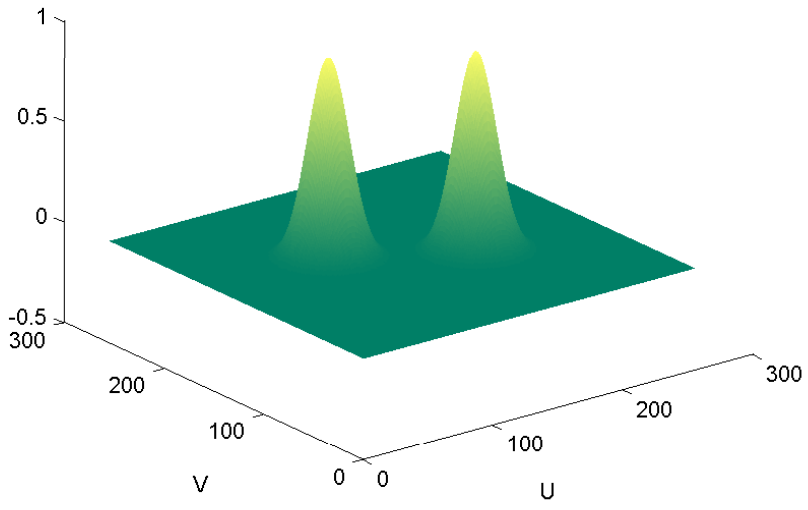


Figure 3-15 the Gabor function in the spatial- frequency domain

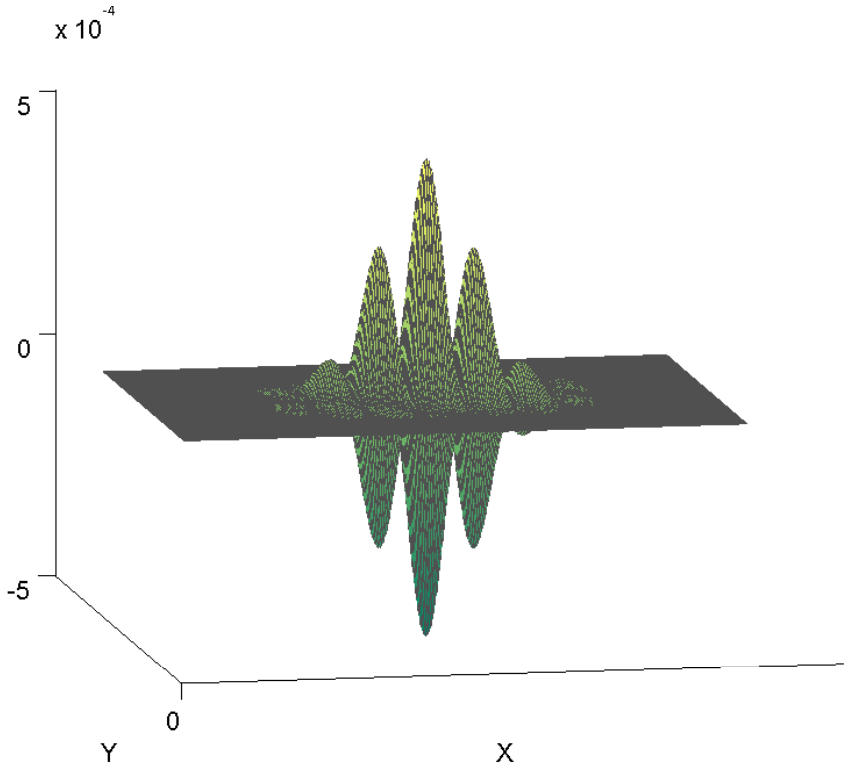


Figure 3-16 The Gabor function in the spatial domain

3.6 Image Filtration in the Frequency Domain

The raw TRUS image as well as the segmented prostate mask and the suspicious regions marked by the radiologist are the inputs for the algorithm. The steps of filtering the image in the frequency domain are as follows:

1. The original TRUS image is read and displayed; a sample of that image is shown in Figure 3-17
2. The mask obtained from the manual segmentation is then used to obtain the segmented prostate.
3. The size of the segmented prostate image is then obtained and is used to determine the filter frequency using equation 3.8 as well as the variables σ_x and σ_y from equation 3.10.
4. The image is transformed to the frequency domain where a real component and an imaginary component of the image are obtained.
5. Both the real and the imaginary components are multiplied by the Gabor function in the frequency domain.

6. Next, some of the feature extraction methods explained in section (3.2.4) are applied to the image real and imaginary components:
    a. As the results of the Brodatz images recommend, the magnitude response is used for the TRUS images. The magnitude response is able to capture fine textures in the Brodatz images and because the TRUS images' textures are usually fine, the magnitude response is used.
    b. The pixel adjacency is suitable for some of the synthesized images as pixels belonging to the same texture are close to each other, therefore they should be clustered in the same cluster. However, this process was not appropriate for the TRUS images and is not expected to perform well as the cancerous region might be small; therefore applying this method will blend the cancerous regions into the neighboring regions.
    c. Gaussian smoothing was also recommended by the Brodatz images' segmentation results, where a Gaussian filter is used to smooth the output of the Gabor filter. The standard deviation of the Gaussian filter is chosen to be the same as that of the corresponding Gabor filter as recommended in [74]. The smoothing obtained better results for the TRUS images as it suppresses the noise that is usually present in the ultrasound images.
7. The contours of the ROIs of the Gabor filtered TRUS images are obtained according to the frequency levels of the output image.
8. A comparison between the original TRUS image, the segmented TRUS image without any filtering, the segmented unsmoothed Gabor filtered TRUS image, and the segmented smoothed Gabor filtered TRUS image is shown in Figure 3-18. It is clear from the figure that the smoothed Gabor filtered image obtained the best segmentation. Therefore, this method is used for the ROI segmentation throughout this thesis.
9. The output of the algorithm stages is shown in Figure 3-19 where (a) represents the input image, (b) represents the magnitude response of the Gabor filtered image, (c) represents the smoothed version of the Gabor filtered image that is then segmented and the segmentation results is shown in (d). It is clear from the figure that the implemented algorithm did capture the cancerous regions as well as some other regions. The suspicious ROIs are then selected among the identified regions by incorporating the doctors' expertise information into the algorithm, which is explained in the next section.

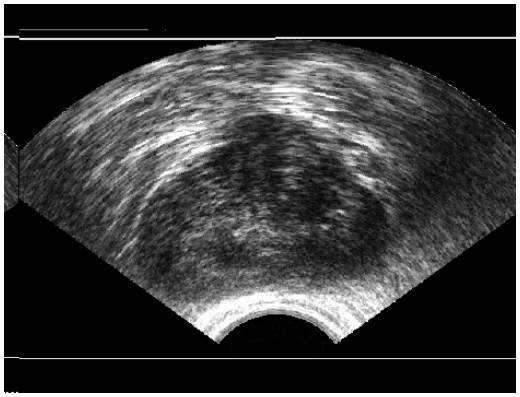


Figure 3-17 The original TRUS Image

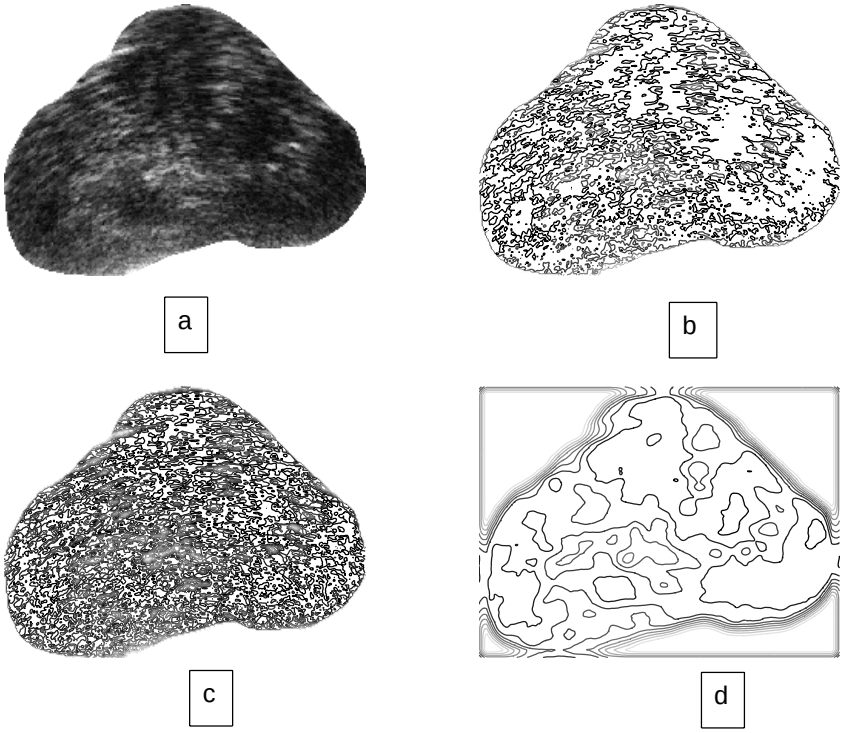


Figure 3-18 a- The original TRUS image, b- The segmented original TRUS image, c- The segmented Gabor filtered image, d- The segmented smoothed Gabor filtered image.

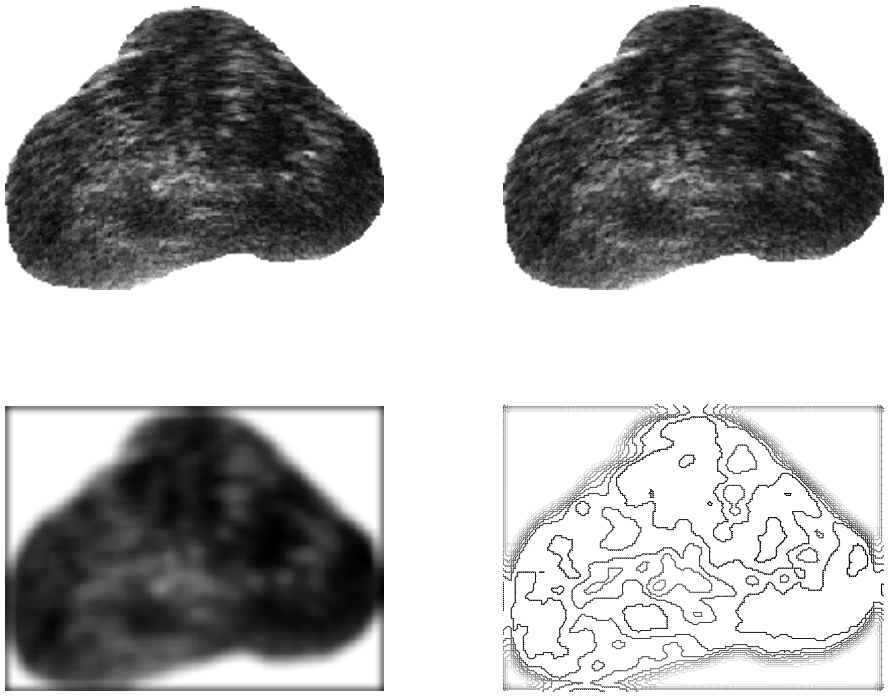


Figure 3-19 Stages implemented to obtain the segmented TRUS image, a- TRUS image, b- Gabor filtered TRUS image, c- smoothed Gabor filtered TRUS image, d- Segmented TRUS image

3.7 Expertise Information Incorporation

The expert radiologist identified regions were examined in this work and their properties were translated into features. This was achieved by the following steps:

1. The available TRUS images are segmented using the proposed algorithm.
2. The segmentation algorithm assigns some features to each segmented regions which are:
    a. The region's contour length
    b. The region's contour location
    c. The region's contour frequency level
3. For each segmented region that matched the radiologist identified regions the above features were recorded.
4. Some of these features are used as criteria for the implemented algorithm. The used features are the contour length and frequency level. The contour location was not used as a criterion

for the algorithm in order not to miss any suspicious regions i.e. to be able to identify the suspicious region regardless of their location.

5. The above explained criteria are integrated into the proposed segmentation algorithm and the algorithm is then applied to all the available TRUS images.

Therefore, the final segmentation algorithm obtains identified regions that have the spatial domain information blended with the frequency domain information as well as the expertise information.

3.8 Results and Discussion

Using the Gabor multi-resolution analysis showed excellent results in recognizing the ROIs selected by the expert radiologist. The radiologist used his intuition to identify these regions; the proposed algorithm did convert the radiologist instinct to features and conditions. The proposed algorithm did use the hidden information extracted from the radiologist's decision and used this information to get to the suspicious regions. Therefore, in some situations the proposed algorithm did identify some regions that were not recognized by the radiologist and satisfy the previously set conditions explained in section (3.7). The obtained results are discussed in the coming sections and figures. The figures show the original TRUS image with the radiologist stained regions to the left, the regions segmented by the Gabor filters in the middle as well as the selected ROIs using the expertise knowledge to the right. The proposed algorithm is capable of identifying the ROIs under different circumstances that are shown in the following subsections. The results that are discussed in section 6.2 also prove the power of the proposed segmentation algorithm.

3.8.1 Typical prostate shape

The proposed algorithm was successful in capturing the suspicious ROIs from the TRUS images with typical prostate shape. Image registration was used earlier in [84] for identifying cancer regions for the typical prostate shape as well as for the atypical prostate shapes. Image registration needs numerous modeling process; moreover, it assigns certain locations in the gland to contain cancerous regions. However, the proposed algorithm didn't have this modeling difficulty and it didn't pre-assign the location constraint to the candidate regions. Therefore, the proposed algorithm is more robust and is able to identify suspicious regions even if these regions are not in a high cancerous probability locations. For the typical gland shape, Figure 3-20, Figure 3-21 and Figure 3-22 show that the algorithm was successful in identifying most of the doctor's identified regions in the typical gland. The proposed algorithm did capture the information that was hidden in the radiologist's decision and can't be seen by the naked eye. In addition, the algorithm also recognized regions that were missed by the radiologist, yet, still carry the same information of the radiologist identified regions which proves the superiority of the proposed algorithm.

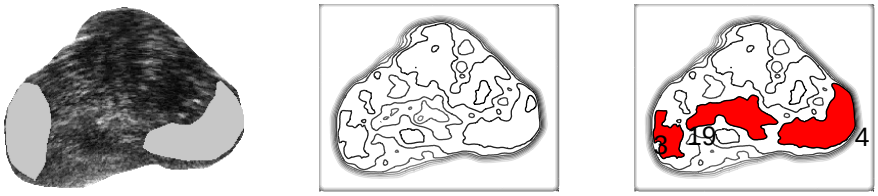


Figure 3-20 typical prostate shape and its corresponding ROIs

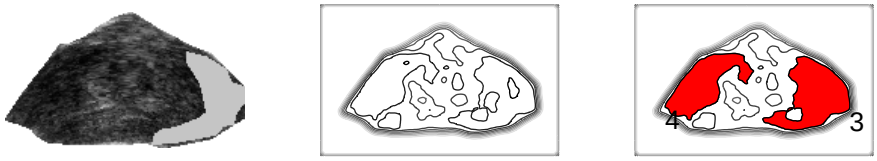


Figure 3-21 typical prostate shape and its corresponding ROIs

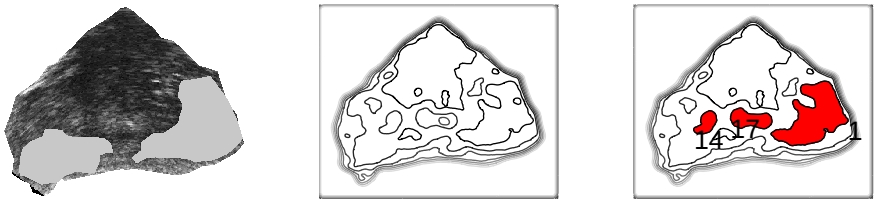


Figure 3-22 typical prostate shape and its corresponding ROIs

3.8.2 Unusual Prostate Shape

The algorithm also shows robustness even if the prostate shape is not the typical walnut shape, a condition in which cancerous regions are most probably lost using the spatial location constraint discussed in section (3.8.1). The efficiency of the proposed algorithm is clear from Figure 3-23, Figure 3-24, Figure 3-25, Figure 3-26 and Figure 3-27 where the regions that were marked by the doctor were recognized in addition to more regions like region number 9 in Figure 3-23 and region number 1 in Figure 3-25. These extra regions did have the same texture features as the regions identified by the radiologist, yet, the radiologist failed to identify them. This result proves the robustness and superiority of the proposed algorithm even for the atypical prostate shape, a case that is sometimes confusing the registration methods. It is also clear from Figure 3-23 and Figure 3-26 that the proposed algorithm did recognize regions that are not located in the highly suspicious cancerous locations, which is an advantage to that algorithm. Moreover, Figure 3-27 shows that the radiologist marked a big area to the top of the prostate image, yet, the proposed algorithm was more precise in identifying the region and extracted it in a finer manner, which will better guide the biopsy operation.

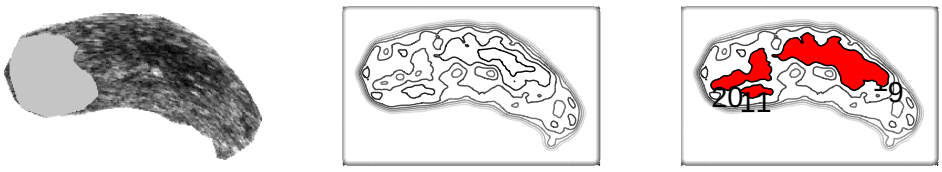


Figure 3-23 Example 1 of ROIs for Atypical Prostate Gland

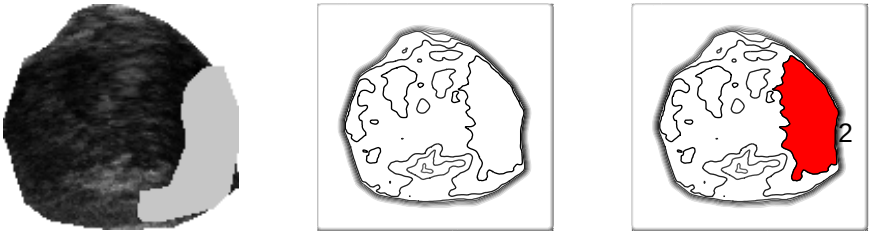


Figure 3-24 Example 2 of ROIs for Atypical Prostate Gland

Figure 3-25 Example 3 of ROIs for Atypical Prostate Gland

Figure 3-26 Example 4 of ROIs for Atypical Prostate Gland

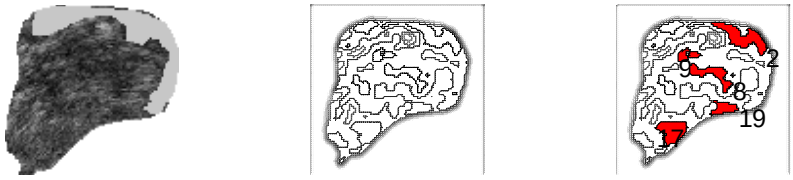


Figure 3-27 Example 4 of ROIs for Atypical Prostate Gland

3.8.3 Recognizing relatively big cancerous regions

It is clear from Figure 3-28, Figure 3-29, Figure 3-30 and Figure 3-31 that the algorithm is not only capable of capturing fine cancerous regions but is also capable of recognizing relatively big cancerous regions regardless of either the prostate shape or the spatial location. This result proves the supremacy the proposed algorithm has over the registration methods and the spatial location contained based methods. The algorithm can even recognize cancerous regions when it occupies the bigger portion of the gland as shown in Figure 3-31 Recognizing relatively big regions . Moreover, these regions are identified in more detailed manner. In fact, the resulting ROIs can be considered as a detailed image that contains much more information than the manually segmented image. Therefore, it could be effective as a decision support tool for radiologists for manual region segmentation.

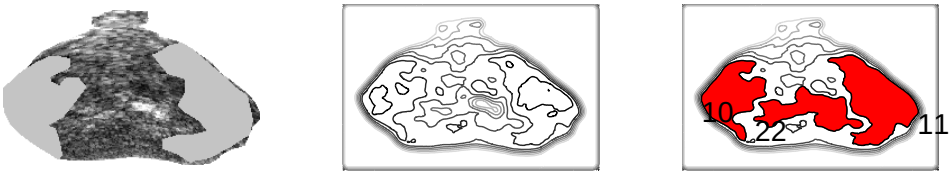


Figure 3-28 Recognizing relatively big regions 1

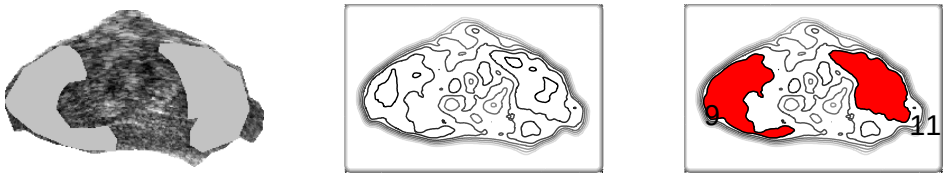


Figure 3-29 Recognizing relatively big regions 2

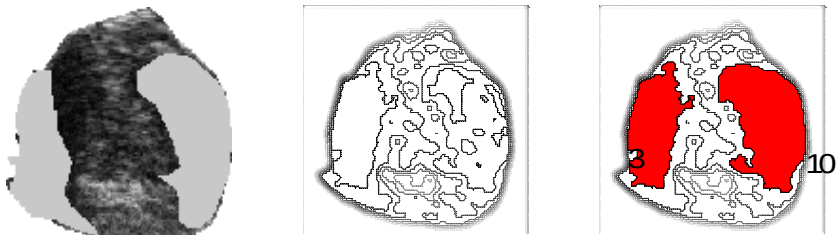


Figure 3-30 Recognizing relatively big regions 3

Figure 3-31 Recognizing relatively big regions 4

3.8.4 Recognizing Relatively Small Cancerous Regions

The proposed algorithm showed excellent recognition even for small cancerous regions and this is clear from Figure 3-32 and Figure 3-33. The proposed algorithm recognized suspicious regions that might have been missed by the radiologist. This proves its ability to follow the pre-set criteria consistently and obtain suspicious regions that were overseen by the radiologist. Moreover, the proposed algorithm is expected be an excellent assistive tool for the radiologist that attracts his attention to these regions and help in the decision making process. Moreover, the algorithm did identify the radiologist identified regions.

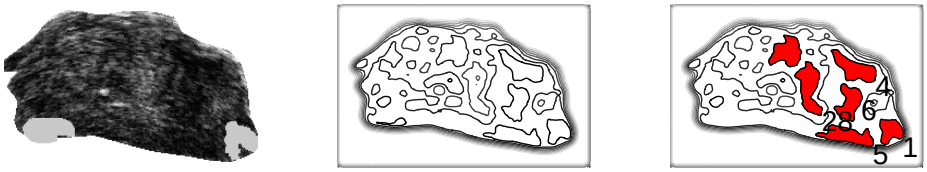


Figure 3-32 TRUS image with small cancerous regions

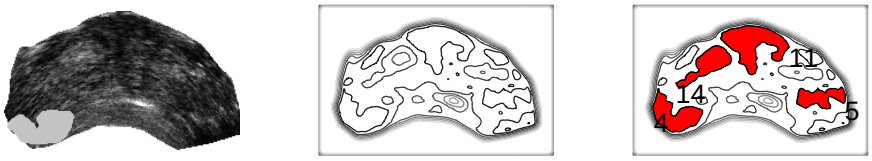


Figure 3-33 TRUS image with small cancerous regions

3.8.5 Recognizing Different Locations for Cancerous Regions

The proposed algorithm is also able to identify suspicious regions regardless of both their location in the TRUS image and the shape of the prostate gland. This can be seen from Figure 3-34, Figure 3-35, Figure 3-36 and Figure 3-37 where the radiologist recognized the bottom region in Figure 3-34, the top region in Figure 3-35 and a region to the right of the TRUS image in Figure 3-36 and the algorithm did recognize all these regions. The algorithm also recognized some isolated regions from the radiologist marked region in Figure 3-37. The segmentation results shown in Figure 3-35 proved the ability of the algorithm to capture regions that would have been missed if a spatial location criteria or a registration method was used. Moreover, the results in Figure 3-37 showed a finer segmentation than the radiologist's segmentation that will facilitate the biopsy needle location choice. Furthermore, the ROI segmentation results shown in Figure 3-36 proved the ability of the algorithm to highlight another suspicious region that was not identified by the radiologist.

Figure 3-34 Cancerous region in the bottom of the TRUS image

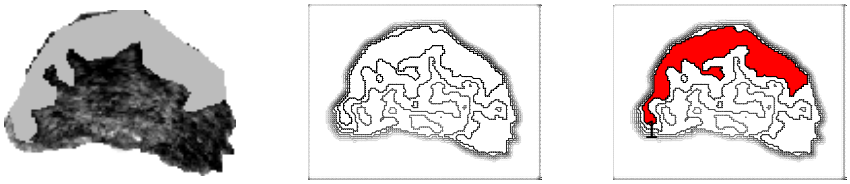


Figure 3-35 Cancerous region in the bottom of the TRUS image

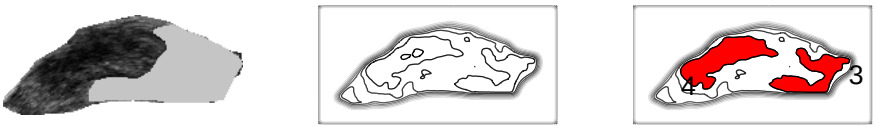


Figure 3-36 Cancerous region in the right of the TRUS image

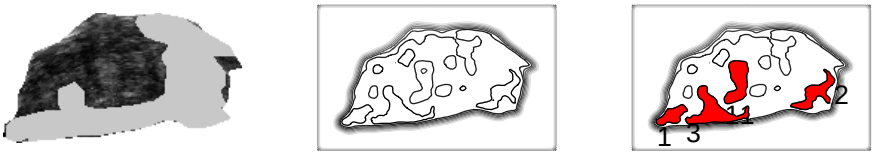


Figure 3-37 Cancerous region in the bottom and right of the TRUS image

In the above sections, it was shown that the proposed ROI segmentation algorithm is superior to different ROI segmentation methods due to the following:

- It is superior to the manual segmentation as radiologists usually correlate the lower echogenic regions in the peripheral zone of the prostate as being cancerous [85]. However, this tumor identification method produces low specificity [86]. On the other hand, the proposed algorithm highlighted regions that follow the same texture pattern of the radiologist identified regions, but are either located in the transition zone or are not darker than the other regions, and therefore were unrecognized by the radiologist.
- The proposed algorithm is also superior to the methods that used image registration as it didn't take into consideration the ROIs location and was able to identify ROIs in any location of the image.
- Moreover, the proposed algorithm saved plenty of time and computational effort as it used only one Gabor filter that is designed specially to fit the TRUS image segmentation problem.
- It also doesn't need to extract computationally demanding textural features from the whole image like the previous methods that divided the image into 18x18 squares and analyzed each square separately [15]. Dividing the image into squares might lead to the study of a square that might have mixed cancer and non-cancer tissues.
- In contrary to the previous algorithms the proposed algorithm doesn't need any training data to learn from. As once the radiologist marked regions' features discussed in section (3.7) are set in the design stage, they can be applied to any TRUS image.
- The proposed algorithm is able to extract finer regions, which will help in the guidance of the biopsy needle during the biopsy operation.

- Unlike the previous methods, such as the approach presented in [70] that requires prior knowledge of the type of tumor (whether hypoechoic, hyperechoic or isoechoic). No prior knowledge of the tumor appearance in the image is required by the proposed algorithm.

3.9 Conclusions

In this chapter a novel ROIs identification method was introduced that depends mainly on the Gabor multiresolusion analysis. This method is composed of several stages. First, the TRUS image was filtered using a high frequency Gabor filter. Next, the magnitude response of the filtered image was smoothed using a Gaussian filter. The following step is to segment the image according to the pixels' frequency level, where the pixels having close frequency level were assigned to the same cluster. And finally the experts' information is incorporated into the algorithm and highly suspicious regions are identified. The ROI identification results are excellent where most of the radiologist marked regions were identified by the algorithm. Furthermore, additional regions were identified that were not selected by the radiologist. Therefore, further analysis of the texture of the identified ROIs is important. Several texture analysis methods were implemented and the analysis results are shown in the coming chapters.

Chapter 4
Statistical Feature Analysis

4.1 Introduction

The output of the previous stage, as discussed in Chapter 3, is some suspicious ROIs that were identified using a novel ROI identification technique. The proposed ROI identification method depends on the response of each pixel to a Gabor filter where pixels that possess a similar response to the filter were assigned to the same region. Moreover, some information was extracted from the doctor's stained regions, on the TRUS images, such as the length of the region's contour and the frequency level of the region. The proposed algorithm did capture the regions that were stained by the doctor in addition to some regions that were not identified by the doctor although these missed regions posses the same hidden textural features. The identified ROIs are further analyzed in the coming chapters in order to complement the proposed recognition algorithm by digging into the ROIs' texture features in order to extract important information from these ROIs. The analysis of the ROIs was achieved using the ROIs image texture features. The feature construction methods used in this chapter already exist in the literature, however, their combination and application to TRUS images is implemented for the first time. Therefore, this chapter is considered an application based research that focuses mainly on the TRUS image characteristics and the features that can enhance the tissue characterization process. It should be noticed at this stage that the gold standards that are available for this research are the radiologist identified images, which makes the obtained results bound by the radiologist's view. Better recognition rates are expected when pathology identified gold standards are available.

Typically, image features can be based on individual pixels (e.g., pixel intensity level or the distance between two pixels with certain grey level), on areas (the detection of regions having specific shapes), on transformations (wavelet, Fourier, etc.) of the original data, or on time (the change in an image since the last examination) [44, 87, 88, 89]. Classification, comparison, or analysis of medical images is performed usually in terms of a set of features constructed from the images. Usually this is necessary for one or more of the following reasons:

1. *Reduction of dimensionality:*

It is worthwhile to express structure within images and similarities between images in ways that depend on fewer, higher-level representations than their pixel values and relationships. It is important to show that the reduction adds information that is important to the task.

2. *Incorporation of cues from human perception:*

Much is known about the effects of basic stimuli on the visual system. Moreover, in many situations, there is a considerable insight into how humans analyze images (essential, for example, in

the training of radiologists). Utilizing the appropriate features allows for the incorporation of that experience into automated analysis.

3. *Transcendence (extension) of the limits of human perception:*

There are many properties (e.g., some textures) of images that cannot be perceived visually, but could be useful in characterizing the images. Features can be constructed from various manipulations of the image such as the higher order statistical features that make those properties evident.

Feature analysis is better to be operator independent in order to get rid of the operator mistakes that affect the recognition consistency depending on different operators and their circumstances. This research is concerned with those features that can be computed without user interaction. These features are constructed and then, in the next chapter, a subset of the constructed features is selected for each of the identified regions of interest (ROIs) in the TRUS images. Feature construction is a process that aims to discover hidden relationships between features (or pixel values and locations), inferring new composite features.

4.2 Statistical Feature Construction Methods

In characterizing texture through statistical approaches, local features are calculated from combinations of intensities at points on a textured surface or image, in our case ROIs. Depending upon the number of points used to form the local features, the statistical measures are classified into first, second, and higher-order statistics. Texture compositions differing in first- and second-order statistics can be discriminated from each other by a human observer. However, those that differ in their third or higher-order statistics, in general, cannot be discriminated by the human observer [90, 91]. Different statistical based texture analysis algorithms are shown in Figure 4-1.

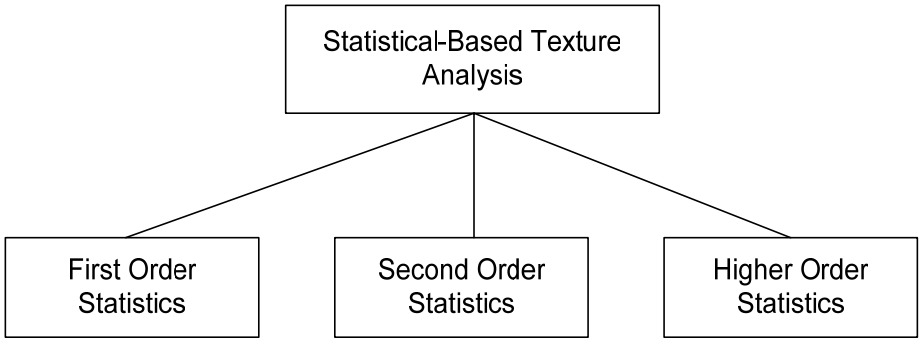


Figure 4-1 Statistical Based Texture Analysis Methods

The most conceptually simple feature construction method is the histogram-based (first order statistics) approach, where textural properties are calculated from the histogram containing the occurring probability of a particular grey-level [44, 87, 88].

In the visual perception of texture, the discrimination of texture from an image depends largely on the difference in the second-order statistics of the image [87]. Measures to detect such differences are based on the joint probability of a pair of grey-levels (i, j) occurring at separation d_s. Calculating this probability measure over an image forms what is known as a co-occurrence matrix. A framework for calculating the co-occurrence matrices was first defined in [92, 93]. However, the construction of features using this technique proved computationally expensive. In an attempt to improve the efficiency, and at the same time retain the general applicability of the scheme, Grey Level Dependence Vector was developed (GLDV) [94]. This technique not only attempts to optimize the feature set, but is also applicable to all kinds of textures.

When the analysis involves the relation between more than two pixels, the features constructed are expected to be more informative than the HVS and this analysis is referred to as higher order statistics.

The TRUS imaging system depends mainly on the ultrasound beam that penetrates through the body tissues; therefore, the image pixels contain information from their neighbors in which the beam has passed through. Consequently, the effect of the neighboring pixels should be accounted for in the proposed analysis. As a result, in this chapter the second and higher order statistics based features are used to extract information from the identified ROIs. First order statistics method is not expected to extract useful information due to the fact that it considers only the intensity level of the pixels without any consideration of the relationship with the neighboring pixels. Therefore according to the previous discussion, it is not applied in this work due to the complexity of TRUS images.

4.3 TRUS ROI Feature Construction

Identified ROIs are subjected to further analysis where statistical features are being constructed from these ROIs. This step is achieved by checking for each considered pixel pair or group of pixels in the image whether it is located inside the specific region under study. The check is repeated for all the identified regions in the image.

It has been shown in [90] that two textures are only differentiable by human eye if they have different second order statistics. However, higher order statistics carry information that is not visible by the human eye and as discussed in section 4.2, there is a relation between each pixel and not only its next neighbor but also the pixels that are not next to it. This is the main reason behind choosing second and higher order statistics features for this specific application. Two different second order statistics feature construction methods, namely; Grey Level Dependence Matrix GLDM or the co-occurrence matrix and Grey Level Difference Vector GLDV were used in this work. Moreover, two higher order statistics feature construction methods were also used in this work namely Neighborhood Grey Tone Difference Matrix (NGTDM) and Grey Tone Run Length Matrix (GTRLM). The details of these methods are explained in the coming sections.

4.3.1 Second-Order Statistics

Second-order statistics describe the spatial properties of the texture. It is considered a very useful characterization that uses subsets of pixels to obtain information on the image texture. Second order statistics is applied in texture analysis of different types of images [66, 95 and 96]. The main criterion in the separation of textures that belong to distinct classes is the difference in their second-order statistics. That is, for two textures having identical second-order statistics a deliberate amount of effort is required by the human visual system to discriminate between them. In contrast, little effort is required when the second-order statistics are different. However, contrary to what one might imagine, textures, which differ in third-order statistics having equal first- and second-order statistics, get away from spontaneous detection by the human visual system. Some different second order statistics algorithms were used in this work for prostate tissue characterization using the ROIs obtained from the TRUS images. These second order statistics methods are explained in the coming sections.

4.3.2 Grey Level Dependence Matrix (GLDM)

Some important texture features can be extracted from the co-occurrence (Gray Level Dependence Matrix GLDM). GLDMs were first introduced by Haralick [92, 93], who extracted some informative features from the GLDM. GLDMs have been found effective in a number of applications such as SAR images [97], Ultrasound images [71] and Terrain classification [98]. GLDMs are matrices whose elements are the probabilities of finding a pixel, which has, grey-tone i at a distance d_s and an angle φ from a pixel, which has grey-tone j. Each pixel is considered as having eight nearest neighbors connected to it, except at the periphery. The neighbors can be grouped into the four categories shown in Figure 4-2 . The texture information is contained in the probability density functions or GLDMs $P(i, j)$.

Where each $P(i, j : d_s : \varphi)$ is the probability with which two neighboring pixels, one with grey-tone i, and the other with grey-tone j occur, when separated by a distance d_s and an angle φ, thus a low contrast image will have a strongly diagonal matrix, with very low counts for matrix elements far from the diagonal elements. If the texture in the image is coarse, d_s is smaller than the texture element's dimension, then pixels separated by d_s will have similar gray levels and there will be many counts along the main diagonal of the matrix. Conversely, high variations within the image (d_s comparable to texture-element size) will appear in the matrix as substantial numbers of counts located far from the diagonal, making the overall matrix uniform. A set of 14 local textural features were generated by Haralick, which are computed in the spatial domain from each of the GLDMs.

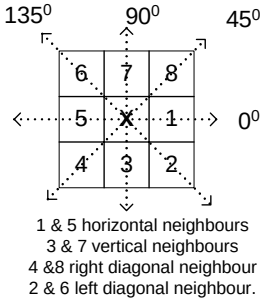


Figure 4-2 Co-occurrence Matrix Demonstration

The features defined by Haralick are:

1. *Angular second moment* (rate of change of shape's area, high values of ASM or Energy occur when the window is very orderly.)
2. *Contrast* "a measure of local image variation"
3. *Correlation* "measures the linear dependency of grey levels on those of neighboring pixels"
4. *Sum of Square Variance*
5. *Inverse Difference Moment or Homogeneity*: "weights values by the inverse of the Contrast weight, with weights decreasing exponentially away from the diagonal"
6. Sum Average
7. *Sum Variance*
8. *Sum Entropy*
9. *Entropy* "an inverse measure of homogeneity i.e. measure of information content"
10. *Difference Variance*
11. *Difference Entropy*
12. *Information Measure on Correlation 1*
13. *Information Measure on Correlation*
14. *Maximal Correlation Coefficient*

One of the most important properties of the co-occurrence approach is being invariant to gray-level transformations. This property is considered a vital aspect for medical imaging applications. The reason is that, although it may not be possible to control the absolute gray-level values from image to image, their relative values are preserved. Thus, descriptors constructed from the co-occurrence matrices of similarly textured images make that similarity evident. This method's features were used in addition to more texture features in [40] to detect and construct useful features from ultrasonic liver images where the average accuracy was about 90%.

In this work not all the features introduced by Haralick were used but a subset that is expected to be shift invariant and is supposed to carry most of the texture information according to the results in [99] was selected. Some GLDM features are found to be shift invariant in [100]; these features are Contrast, Entropy, Energy, Correlation and Inverse Difference Moment (Homogeneity). Moreover, some of those features were used earlier for fetal lung maturity [101] and liver tissue classification from ultrasound images, which yields to good results [40]. Therefore, the features among the ones recommended in [100] that appear to carry more information about the texture and are not expected to be highly correlated as recommended in [102] were used in this work. Moreover, since one expects that there are more grey level variations in the cancerous regions than that in the non-cancerous regions, then, Contrast, Entropy, Homogeneity (IDM) and Energy (ASM) of a cancerous region are expected to be different and distinguishable than those of a non-cancerous region. From the above discussion it is concluded that the features used in this work are:

1. *Contrast* which is a measure of local image variation, a contrast measure result in larger numbers for more contrasty windows.

$$Contrast = \sum_{i=1}^{N_q-1} n^2 \left[\sum_{i=1}^{N_q} \sum_{j=1}^{N_q} p'(i,j) \right]_{|i-j|=n} \quad \textbf{4-1}$$

2. *Entropy* "a measure of information content"

$$Entropy = -\sum_{i=1}^{N_q} \sum_{j=1}^{N_q} p'(i,j) \log\{p'(i,j)\} \quad \textbf{4-2}$$

3. *Energy*, also named as Uniformity, use each P_{ij} as a weight for itself. High values of Energy occur when the window is uniform. Due to the non-uniformity of the cancerous regions, the energy feature is expected to be powerful in distinguishing cancerous regions.

$$ASM = \sum_{i=1}^{N_q} \sum_{j=1}^{N_q} p'(i,j)^2 \quad \textbf{4-3}$$

4. *Homogeneity*: There are more grey level variations in the cancerous regions than that in the non-cancerous regions. Therefore, homogeneity should posses high discriminatory power and is also used among the GLDM feature set extracted in this work

$$IDM = \sum_{i=1}^{N_q} \sum_{j=1}^{N_q} \frac{1}{1+(i-j)^2} p'(i,j) \quad \textbf{4-4}$$

Where,

$P(i,j)$ represents GLDM.

R is the number of pixels in a given ROI (area).

N_q is the number of distinct grey-tones in each ROI.

$p'(i,j)$ Represent the $(i,j)^{th}$ entry in the normalized GLDM $p'(i,j) = \frac{P(i,j)}{R}$.

There are 108 identified ROIs from 33 TRUS images provided by the medical team in University of Western Ontario. Ninety regions were used for training the classifier while 18 were used for testing the features discriminatory power. These four GLDM features were constructed from each identified ROI and a GLDM feature set is obtained. The results of using the GLDM feature set are discussed in section 4.7.

4.3.3 Grey Level Difference Vector (GLDV)

Co-occurrence methods yield good results in texture discrimination, but are computationally expensive. A variety of modifications and computational strategies have been proposed to reduce the cost. In this work, the Grey Level Difference Vector (GLDV) was used as a short form of the GLDM method. This method is expected to retain the important information present in the GLDM method.

The description of GLDV is given in [40] where it was among the features used for liver tissue classification and it is summarized as follows: Let $I(x,y)$ be the image intensity function. For any given displacement $\delta \equiv (\Delta x, \Delta y)$ let $I_\delta(x,y) \equiv |I(x,y) - I(x+\Delta x),(y+\Delta y)|$ and p_δ be the probability density of $I_\delta(x,y)$. If there are m grey levels, this has the form of m-dimensional vector whose i^{th} component is the probability that $I_\delta(x,y)$ will have the value i. It is simple to compute p_δ by counting the number of times each value of $I_\delta(x,y)$ occurs where $\Delta x \; and \; \Delta y$ are integers. If the texture is course, and δ is small compared to the texture element size, the pairs of points at separation δ should usually have similar grey levels, so that $I_\delta(x,y)$ should usually be small, i.e. the values in p_δ should be concentrated near $i \approx 0$. On the other hand for fine texture the values in p_δ should be more spread out. The GLDV feature set is used in [101] where, the feature used was the entropy of p_δ.

In this work, the GLDV was used as a texture measure and the same features explained in the previous section are constructed. The constructed features are expected to carry texture information that is adequate to the characteristics of TRUS images; moreover, these features are not obvious to be highly correlated. The constructed features are:

1. Contrast

$$Contrast = \sum_i i^2 p(i \mid d) \qquad \textbf{4-5}$$

2. Entropy

$$Entropy = \sum_i p(i \mid d) \log\big(p(i \mid d)\big) \quad \textbf{4-6}$$

3. Homogeneity: Inverse Difference Moment (a measure for similarity, it is expected to be useful as cancerous regions texture is considered non-similar)

$$IDM = \sum_i \frac{p(i \mid d)}{\left(i^2 + 1\right)} \quad \textbf{4-7}$$

4. Energy: Angular Second Moment is a measure of uniformity which is supposed to be different for cancerous regions than that of the non-cancerous regions as the cancerous regions are known to be highly non-uniform.

$$ASM = \sum_i \big(p(i \mid d)\big)^2 \quad \textbf{4-8}$$

5. Mean: it was recommended in [94, 103] to use the Mean in addition to the four previous features for the case of GLDV features in order to enhance texture discrimination.

$$Mean = \sum_i i p(i \mid d) \quad \textbf{4-9}$$

The Energy, Homogeneity, Entropy, Contrast as well as the Mean were constructed from the GLDV and the classification results are shown in section 4.7 .

4.4 Higher-Order Statistics

With the relative infancy of CAD based on TRUS imaging and its' feature analysis as well as its' contribution to the diagnosis of prostate cancer, it would be inappropriate to consider only analysis of the images using second order statistical textural analysis techniques. Moreover, as textures differing in third- or higher-order statistics seem to surpass the capabilities of the human perceptual system, it is appropriate to consider the higher-order analysis of TRUS information. To the best of the author's knowledge, the higher order statistical features were not used earlier in any research involving TRUS images. A statistics is called higher order if it involves three or more pixels. Two different higher order statistics methods were used in this work for the first time in the context of TRUS image feature analysis. These feature construction methods are namely Neighborhood Grey Tone Difference Matrix (NGTDM) and Grey Tone Run Length Matrix (GTRLM).

4.4.1 Neighborhood Grey-Tone Difference Matrix (NGTDM)

In order to benefit from studying the higher order statistics of the ROIs textural properties, the NGTDM method was used in this work. NGTDM developed in [104] attempts to mimic the mechanism with which the human visual system discriminates texture. It was built on the basis that the properties, which humans use to discriminate between different textural patterns, include Coarseness, Contrast, Complexity, Busyness (fineness), Shape, Directionality and Texture strength.

In this approach, the i^{th} entry in a NGTDM is a summation of the differences between all pixels with grey-tone i and the average value of their surrounding neighbors. If i is the grey-level at (x, y) then the average grey-level over a neighborhood centered at (x, y) is:

$$A_i = A(x, y) = \frac{1}{W-1} \sum_{m=-d}^{d} \sum_{n=-d}^{d} i(x+n, y+m) \qquad \textbf{4-10}$$

Where $(m, n) \neq (0,0)$, d specifies the neighborhood size and $W = (2d+1)^2$ it follows that the i^{th} entry in the NGTDM is given as:

$$s(i) = \begin{cases} \sum |i - A_i|, & for\ i \in N_i\ if\ \ N_i \neq 0 \\ 0 & otherwise \end{cases} \qquad \textbf{4-11}$$

Where $\{N\}$ is the set of all pixels having grey-tone i. From the NGTDM, computational approximations were derived for the textural attributes previously mentioned. NGTDM attempts to optimize the task of mathematical feature construction through establishing features, which are connected to the way in which the human visual system psychologically perceives texture primitives. In this work, a square of 3x3 was chosen as the neighborhood where each point in the square should be inside the ROI to avoid taking into account any non-suspicious pixels.

Coarseness: almost the most fundamental psychological visual cue, coarseness refers to the density of edge elements; the finer the texture, the higher the edge density. In a coarse texture grey-tone differences are small indicating that the primitives (building blocks) making up the texture are large. The coarseness measure is obtained from the reciprocal of the summation of intensities at all points in relation to the intensity of the surrounding neighbors, However in the summation each entry is weighted by the probability of occurrence of the corresponding intensity value p_i. This feature is adequate to the TRUS ROIs as the coarseness of the cancerous regions is supposed to be different from the coarseness of non-cancerous regions.

$$Coarseness = \left[\varepsilon + \sum_{i=0}^{N_g - 1} p_i(s_i) \right]^{-1} \qquad \textbf{4-12}$$

Contrast: This conceptually simple property enables the distinction between texture classes when clearly different intensity levels are visible. Probably the most influential property of contrast is the dynamic range of grey-levels in an image. If it is large, it usually corresponds to a high difference in intensity and consequently a large value of contrast. In addition, the period at which the pattern repeats, or its spatial frequency, can be quantified by contrast. Contrast is another feature constructed from the NGTDM.

$$Contrast = \left[\frac{1}{N_d(N_d - 1)} \sum_{i=0}^{N_g - 1} \sum_{j=0}^{N_g - 1} p_i p_j (i-j)^2 \right] \left[\frac{1}{n^2} \sum_{i=0}^{N_g - 1} s(i) \right] \quad \textbf{4-13}$$

Where p_i and p_j are the probabilities of having a grey level i and j respectively and N_d is the total number of grey levels present in the ROI.

$$N_d = \left[\sum_{i=0}^{N_g - 1} Q_i \right] \quad where \quad \begin{Bmatrix} Q_i = 1 & if \;\; p_i \neq 0 \\ 0 & otherwise \end{Bmatrix} \quad \textbf{4-14}$$

Busyness: A busy texture is a texture in which there are quick changes of intensity from one pixel to its neighbor. In other words, the spatial rate of intensity changes is very high. If these changes are small in magnitude, they may not be visually obvious and a high level of local uniformity in intensity may be apparent. On the other hand, if the spatial rate of changes in intensity is low, a high degree of local uniformity may still be perceived, even if the magnitude of the changes is large. While the spatial rate of intensity changes reflects the level of busyness, the magnitude of these changes depends upon the dynamic range of gray scale, and thus relates to contrast. Therefore this feature fits as a measure for TRUS ROIs texture as cancerous regions are expected to be busier (have more rapid changes) than their non-cancerous counterparts.

$$Busyness = \left[\sum_{i=0}^{N_g - 1} p_i s(i) \right] \left[\sum_{i=0}^{N_g - 1} \sum_{j=0}^{N_g - 1} i p_i - j p_j \right]_{p_i \neq 0,\, p_j \neq 0}^{-1} \quad \textbf{4-15}$$

Complexity: This textural property is connected with the amount of information that is present within a textured object. A texture having large number of primitives, each having varying average intensities, is said to be a complex texture. Therefore, clearly for a complex texture having many primitives, each bounded by edges, the properties of contrast and busyness could be used to quantify the texture. Thus, the measure of complexity is partially correlated with busyness and contrast. It is clear from that definition that the complexity is expected to be an informative feature for TRUS ROIs characteristics as the cancerous regions are expected to be more complex.

$$Complexity = \sum_{i=0}^{N_g - 1} \sum_{j=0}^{N_g - 1} \left\{ \frac{\left(|i-j| \right)}{\left(n^2 \left(p_i + p_j \right) \right)} \right\} \left\{ p_i s(i) + p_j s(j) \right\}_{p_i \neq 0,\, p_j \neq 0} \quad \textbf{4-16}$$

Texture strength: Textures, which have a strong impact on the psychological perception are said to possess textural strength. Such textures tend to have well-defined primitives and any measure of texture strength is dependent upon the distinction between these primitives. This suggests that there is a correlation between the coarseness and contrast to the strength of a texture. Texture strength is also used among the texture features that characterize the TRUS ROIs in this work. This is due to the

expectancy that texture possesses more strength in cancerous regions than its strength for non-cancerous regions.

$$Strength = \left[\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (p_i + p_j)(i-j)^2 \right] \left[\varepsilon + \sum_{i=0}^{N_g} s(i) \right]^{-1}_{p_i \neq 0, p_j \neq 0} , \quad \textbf{4-17}$$

p_i & $p_j \neq 0$ to guarantee positive definite features' values

4.4.2 Grey Tone Run Length Matrix (GTRLM)

Another higher order statistics feature construction method which is the Grey Tone Run Length Matrix (GTRLM) was applied for the TRUS images ROIs. This method is based upon the analysis of higher-order information content. GTRLMs [105] compute the run length of a particular image grey-tone in a direction α within a textured image. A run is a set of consecutive pixels having the same or similar intensities along a specified direction. The pixels considered must be linearly adjacent in the specified direction of α. The number of pixels contained within the run is the run length. For example, one may expect that in a coarse texture there are relatively long runs, which occur regularly whereas for a fine texture one would expect to find much shorter runs. The number of runs with grey-tone i of run length j in some α direction is denoted by:

$$R(\alpha) = \left[r'(i, j/\alpha \right] \quad \textbf{4-18}$$

The gray level difference considered in this research is set to 1 which means that the run stops whenever the grey level changes by more than 1. Five textural features are computed from the matrices as recommended in [105]:

Short Run Emphasis

$$SRE = \frac{1}{T_R} \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_r} \frac{r'(i, j|\alpha)}{j^2} \quad \textbf{4-19}$$

Long Run Emphasis

$$LRE = \frac{1}{T_R} \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_r} j^2 r'(i, j|\alpha) \quad \textbf{4-20}$$

Grey-Tone Distribution

$$GTD = \frac{1}{T_R} \sum_{i=0}^{N_g - 1} \left[\sum_{j=1}^{N_r} r'(i, j|\alpha) \right]^2 \quad \textbf{4-21}$$

Run Length Distribution

$$RLD = \frac{1}{T_R} \sum_{j=1}^{N_r} \left[\sum_{i=0}^{N_g - 1} r'(i, j|\alpha) \right]^2 \quad \textbf{4-22}$$

Run Percentages

$$RP = \frac{1}{T_P} \sum_{j=1}^{N_r} \left[\sum_{i=0}^{N_g - 1} r'(i, j|\alpha) \right]^2 \quad \textbf{4-23}$$

Where, N_g is the maximum number of possible grey-tones, N_r is the number of run lengths in the matrix, and,

$$T_R = \sum_{i=0}^{N_g - 1} \sum_{j=1}^{N_r} \left[r'(i, j|\alpha) \right] \quad \textbf{4-24}$$

T_P is the number of points in the ROI.

4.5 ROI Texture Features' Statistical Analysis

The work done in the previous sections of this chapter is mainly constructing different statistical features. In this section, some of the different constructed features are statistically analyzed, where their histograms, box plots as well as their corresponding statistical measures are presented for both data classes (the cancerous regions and the non-cancerous regions).

The statistical measures as well as the features' histograms and box plots dig into the features' quality and show an indication of the information each feature possess regarding the tissue characterization ability. In other words, this analysis highlights each feature's ability to reflect the tissue type.

From the classification results that are presented in this chapter in section 4.7, the second order statistics features are recommended at this stage as they obtained better classification results compared to their higher order statistics counterparts. The statistical measures that were used in this

work characterize and analyze each of the second order statistics features. These statistical measures are:

1- Mean: The mean is a measure of where the center of the distribution lies. It is the sum of all observations divided by the number of observations. The mean is strongly influenced by extreme values (outliers).

2- Median: The median (also called the 2nd quartile) is the middle observation in the data set. It is determined by ranking the data and finding observation number [N + 1] / 2. The median is less sensitive to extreme values than the mean therefore, it is usually recommended for data with outliers or skewed data.

3- The Standard Deviation (StD): is a measure of how far the observations diverge from the mean. It is equivalent to an average distance from the mean. The standard deviation is the most commonly reported measure of scattering. Similar to the mean the StD is sensitive to outliers. The StD is calculated by first calculating the deviation of each data point from the mean, adding the squares of these deviations which obtain the variance and then the StD is the square root of the variance.

$$StD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} \qquad \textbf{4-25}$$

4- The first quartile Q1: (also called the 25th percentile) a number for which data of 25% of the observations is less or equal to that number. The first quartile is the same as the median of the data, which are less than the overall median. It is calculated by ranking the observations, Q1 is the value of observation number (N + 1) / 4

5- The third quartile Q3: (also called the 75th percentile) a number for which data of 75% of the observations is less than or equal to that number. The third quartile is the same as the median of the part of the data, which is greater than the median. It is calculated by ranking the observations, Q1 is the value of observation number 3*(N + 1) / 4

6- Skewness: is a measure of asymmetry (the tail of the data on one side is longer than the tail on the other side). A negative value indicates skewness to the left, and a positive value indicates skewness to the right.

$$Skewness = \frac{N}{(N-1)(N-2)} \sum \left[(x_i - \bar{x}) / s \right]^3 \qquad \textbf{4-26}$$

7- Kurtosis is the degree of peakedness of a distribution. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. A positive value typically indicates that the distribution has a sharper peak, thinner shoulders, and fatter tails than the normal distribution. A negative value means that a distribution has a flatter peak, fatter shoulders, and thinner tails than the normal distribution

$$\text{Kurtosis} = \frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum\left[(x_i - \bar{x})/s\right]^4 - \frac{3N(N-1)^2}{(N-2)(N-3)} \qquad \textbf{4-27}$$

Where N is the number of observations and s is the StD.

The defined statistical measures were calculated for the cancerous regions as well as the non-cancerous regions using the MINITAB statistics software [116] and the results are shown in Table 4-1.

Table 4-1 Statistical Measures of the Constructed Features

		Mean	**StD**	**Q1**	**Median**	**Q3**	**Skewness**	**Kurtosis**
Cancer	GLDV	**114**	**131**	**26**	**45**	**176**	**1.3**	**0.4**
Non-Cancer	**Contrast**	96.1	113	16	72	129	2	6
Cancer	GLDV	**6.5**	**3.6**	**4**	**5.2**	**8.2**	**1.2**	**0.5**
Non-Cancer	**Mean Avg**.	5.8	3.6	3.4	4.8	7.4	2	5.8
Cancer	GLDV	**2.15**	**0.5**	**1.86**	**2.23**	**2.5**	**-0.24**	**-0.25**
Non-Cancer	**Entropy**	1.97	0.62	1.5	2	2.3	-0.6	0
Cancer	GLDV	**0.17**	**0.08**	**0.11**	**0.141**	**0.23**	**1.05**	**0.5**
Non-Cancer	**IDM**	0.2	0.1	0.1366	0.168	0.235	1.16	0.88
Cancer	GLDV	**0.19**	**0.12**	**0.11**	**0.14**	**0.22**	**1.69**	**2.46**
Non-Cancer	**ASM**	0.24	0.18	0.12	0.18	0.27	2.02	3.9
Cancer	GLDM	**4.8**	**1.8**	**4**	**5.15**	**6**	**-0.86**	**-0.15**
Non-Cancer	**Contrast**	4.4	1.6	3.7	4.7	5.7	-1.05	0.12
Cancer	GLDM	**0.08**	**0.15**	**0.003**	**0.007**	**0.06**	**1.94**	**2.47**
Non-Cancer	**Energy**	0.12	0.21	0.006	0.013	0.07	2.23	4.11
Cancer	GLDM	**102**	**124**	**18**	**37.5**	**160**	**1.36**	**0.51**
Non-Cancer	**Entropy**	85.5	106.4	10	61.7	117	2.31	6.29
Cancer	GLDM	**0.42**	**0.14**	**0.32**	**0.37**	**0.54**	**0.9**	**0.7**
Non-Cancer	**Homog**.	0.47	0.17	0.37	0.42	0.53	0.96	0.91

These statistics are also shown in the figures that represent the feature histogram and its corresponding normal fit. The histogram of the data overlaid with a normal curve is usually used to assess the data normality.

Furthermore, the **Anderson Darling normality test** was performed on the constructed features and the results are shown in each figure. The measures obtained from this test are the *A Squared* and the *p* values that are considered other measures of whether the data follows a normal distribution or not.

- The ***A Squared*** measures the area between the fitted normal distribution and the plot points, the smaller the *A squared* value indicates that the normal distribution fits the data better.
- The ***p*** value was calculated from the *A Squared* value. If the *p* value is less than 0.005 this means that the data don't follow normal distribution according to the Anderson Darling null hypothesis.

The Anderson-Darling normality test is defined by a null hypothesis where:

H_0: The data follow a normal distribution (null hypothesis).

H_A: The data do not follow a normal distribution (alternative hypothesis).

If the p value is less than 0.005 means the null hypothesis (H_0) is rejected in favour to the alternative hypothesis (H_1)

These measures are explained in details in [115].
The figures from Figure 4-3 to Figure 4-14 show different features' histograms and their corresponding normal fits. The figures also have the statistical measures corresponding to each feature shown as a legend in the same figure.

Moreover, each figure contains a Boxplot that is located under the histogram. Boxplots summarize information about the shape, dispersion, and center of the data as follows:

- The left edge of the box represents the first quartile (Q1), while the right edge represents the third quartile (Q3). Thus, the box portion of the plot represents the interquartile range (IQR), or the middle 50% of the observations.
- The line drawn through the box represents the median of the data.
- The lines extending from the box are called whiskers. The whiskers expand outwards to point to the lowest and highest values in the data set.

The boxplot can be used to assess the symmetry of the data:

- If the data are rather symmetric, the median line will be approximately in the middle of the IQR box and the whiskers will be similar in length.
- If the data are somewhat skewed, the median may not fall in the middle of the IQR box, and one whisker will likely be noticeably longer than the other.
- This symmetry assessment might not be that accurate in case of the presence of outliers [116]

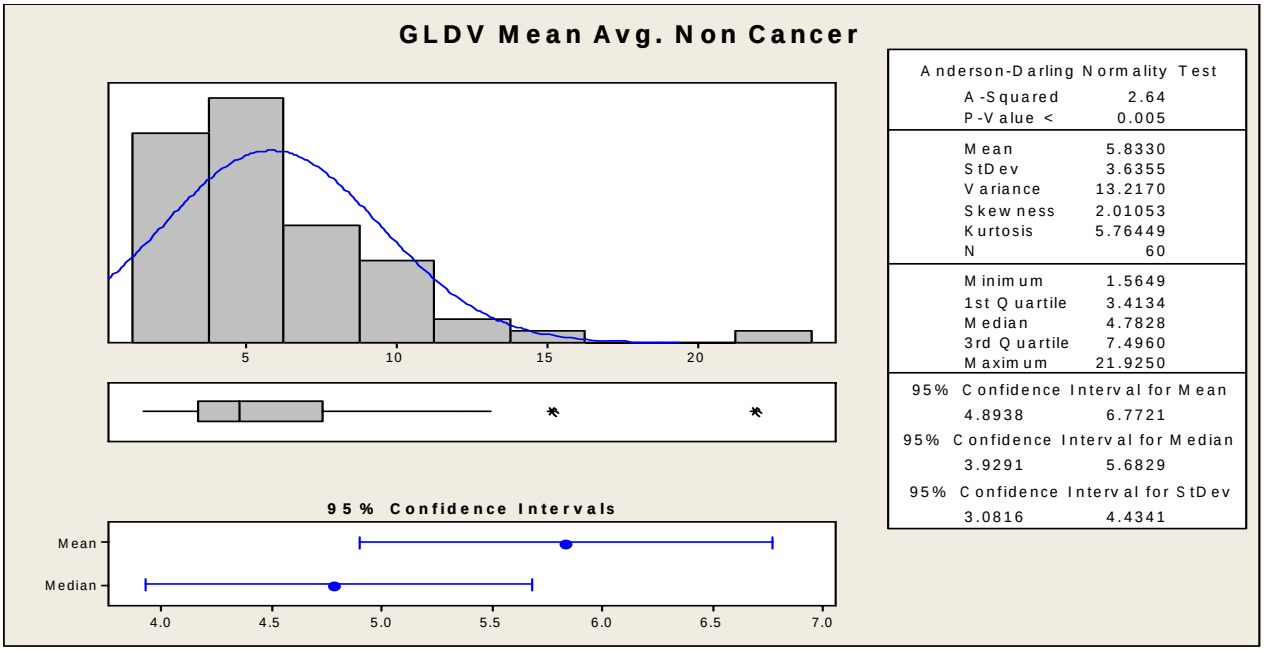


Figure 4-3 GLDV Mean Avg. Histogram and normal fit for Non-Cancer Regions

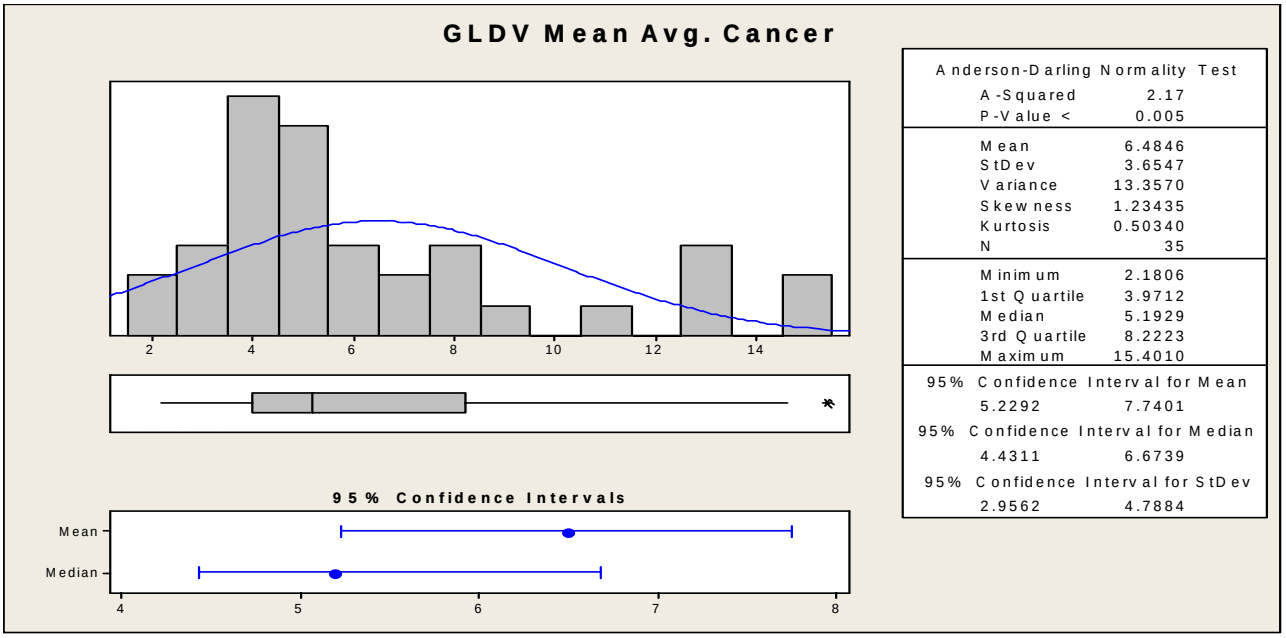


Figure 4-4 GLDV Mean Avg. Histogram and normal fit for Cancer Regions

Figure 4-3 and Figure 4-4 show the Mean Average feature for the non-cancer and cancer regions, respectively. There is an obvious difference between the histogram plot as well as the box plot of both cases. Moreover, the statistical measures of these two features are different where the non-cancer histogram and is more skewed than the cancer one due to the presence of outliers. This is clear from the skewness measure as well as the box plot that showed the outliers. Moreover, the cancer regions possess *Kurtosis* that is much smaller than that of the non-cancer regions, an observation that highlights the fact that the cancer regions histogram is closer to the normal distribution. This result agrees with the A-Squared values in which it is smaller for the cancer regions.

Studying these figures illustrates the difference between the statistics of the Mean Avg. feature of the cancer and non-cancer regions i.e. there is a piece of information regarding the class in the this feature.

The same analysis was performed for the Contrast feature. The histogram of the non-cancer regions and the cancer regions contrasts are shown in Figure 4-5 and Figure 4-6, respectively. From the figures, it is clear that the histogram of the cancer regions is more flat than that of the non-cancer regions. Moreover, the Kurtosis of the cancer regions is smaller than that of the non-cancer regions, which indicates a closer to normal distribution. This is supported by the A-Squared value, which is smaller in case of cancer regions. Furthermore, the boxplots of both histograms are different where the median falls almost in the middle of the plot in case of the non-cancer region and it is shifted to the left in case of the cancer regions, which indicates that the cancer regions are less symmetric. However, the Skewenes value didn't indicate that because of the presence of outliers. From the figures and the statistical analysis it can be concluded that the contrast features adds another piece of information and have some discriminatory power between the cancer and non-cancer regions.

Figure 4-7 and Figure 4-8 show the ASM feature for the non-cancer and cancer regions, respectively. From the A squared values, it is clear that the cancer histogram is closer to the normal distribution than the non-cancer histogram which is supported by the kurtosis value that is smaller in the cancer histogram. From the figure, it is clear that the ASM histogram is concentrated at different values than the non-cancer histogram. This shows that the ASM value is informative for differentiating between cancer and non-cancer regions. However, it might be more informative when combined with the other features.

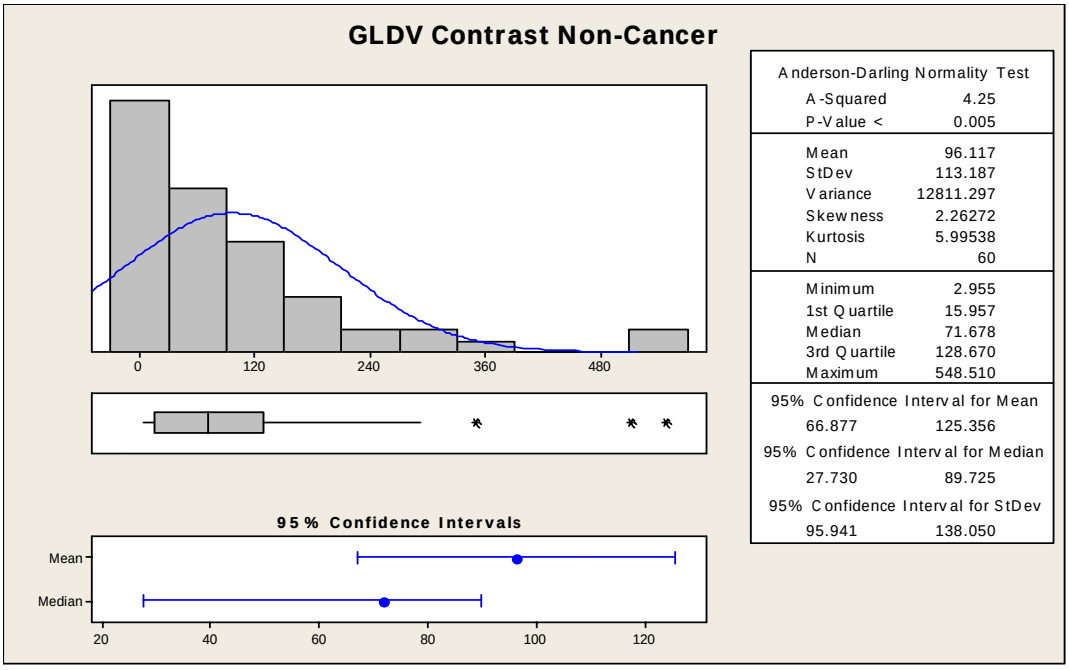


Figure 4-5 GLDV Contrast Histogram and normal fit for Non-Cancer Regions

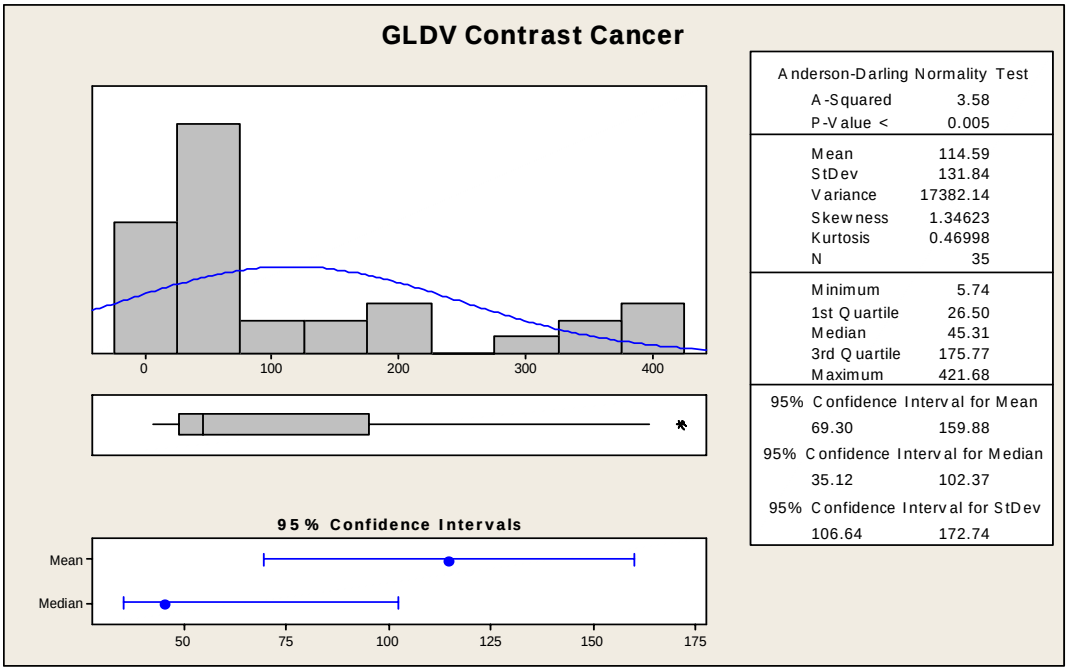


Figure 4-6 GLDV Contrast Histogram and normal fit for Cancer Regions

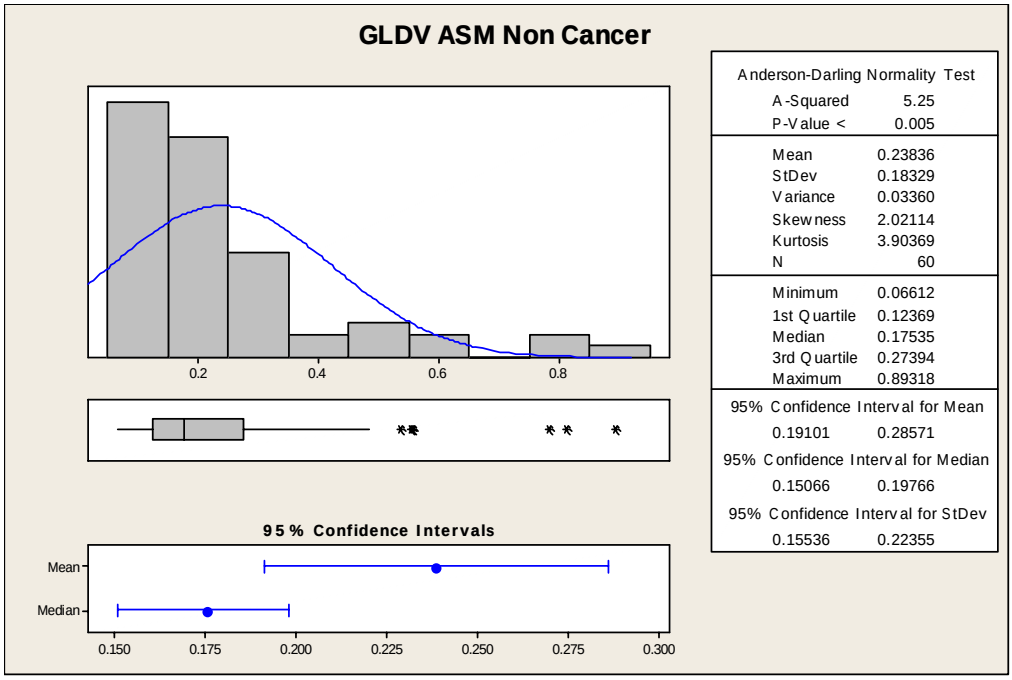


Figure 4-7 GLDV ASM Histogram and normal fit for Non-Cancer Regions

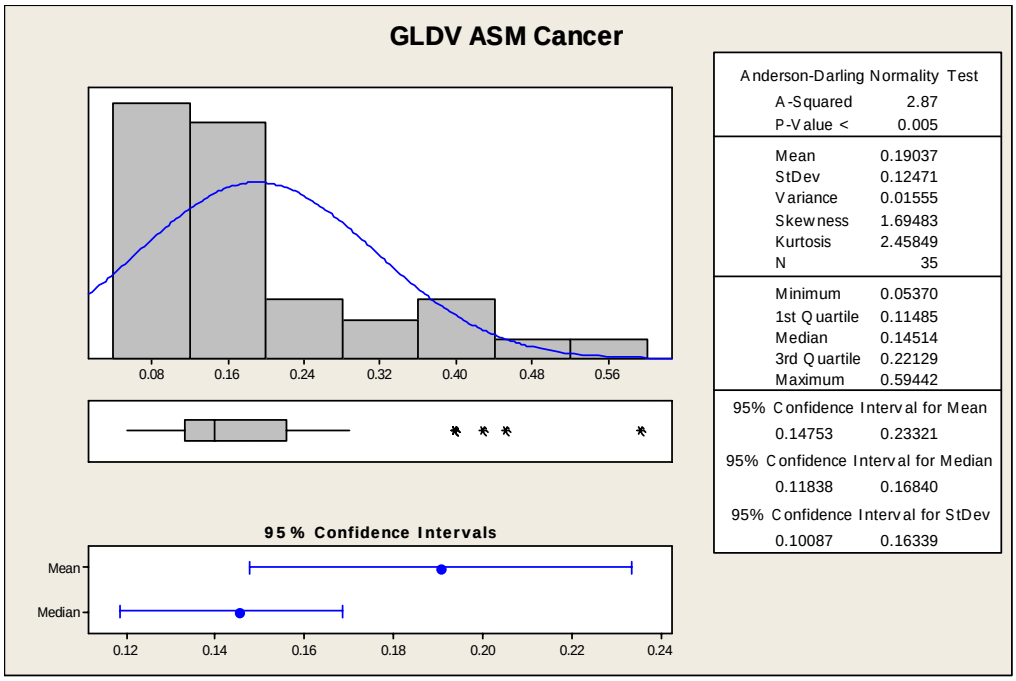


Figure 4-8 GLDV ASM Histogram and normal fit for Cancer Regions

Figure 4-9 and Figure 4-10 illustrate the IDM feature for both the non-cancer and the cancer regions, respectively. There is a visual difference between both figures; moreover, the statistical measures obtained are quite different. The A-Squared value for the non-cancer regions is larger than that of the cancer regions which is consistent with the kurtosis values that is also smaller in the case of the cancer regions. Furthermore, the boxplots of both figures are different where the boxplot of the cancer regions looks more skewed than the non-cancer regions, yet the skeweness didn't indicate this difference because of the presence of the outliers in the non-cancer regions data. This discussion shows that the cancer regions IDM histogram is closer to the normal distribution than its non-cancer counterpart is. From the figure and the statistical measures, the IDM can provide some information about the regions' class.

Figure 4-11and Figure 4-12 portray the entropy feature for the non-cancer and cancer regions, respectively. From the figures, it is clear that the histogram of the non-cancer regions is flatter than that of the cancer regions. Moreover, the A squared value of the non-cancer regions is smaller than that of the cancer regions which indicates that the non-cancer regions histogram is closer to the normal distribution than the cancer regions. However, the p value for both histograms is larger than 0.005 which accepts the null hypothesis that they follow normal distribution.

The above discussed figures are all generated using the second order statistics mainly the GLDV features, which obtained better classification results as will be shown in the next section. However, some of the GLDM cancer regions features' statistics and histograms are different from their non-cancer counterparts such as the GLDM homogeneity feature. The GLDM homogeneity feature for both non-cancer and cancer regions are shown in Figure 4-13 and Figure 4-14, where the histogram as well as the statistical measures are different in both figures. This proves that all the constructed features are considered useful and each has a bit of information that supports decision making. The statistical measures were also summarized in Table 4-1 where all the previously discussed statistical measures for the second order features are shown for both the cancer and non- cancer regions.

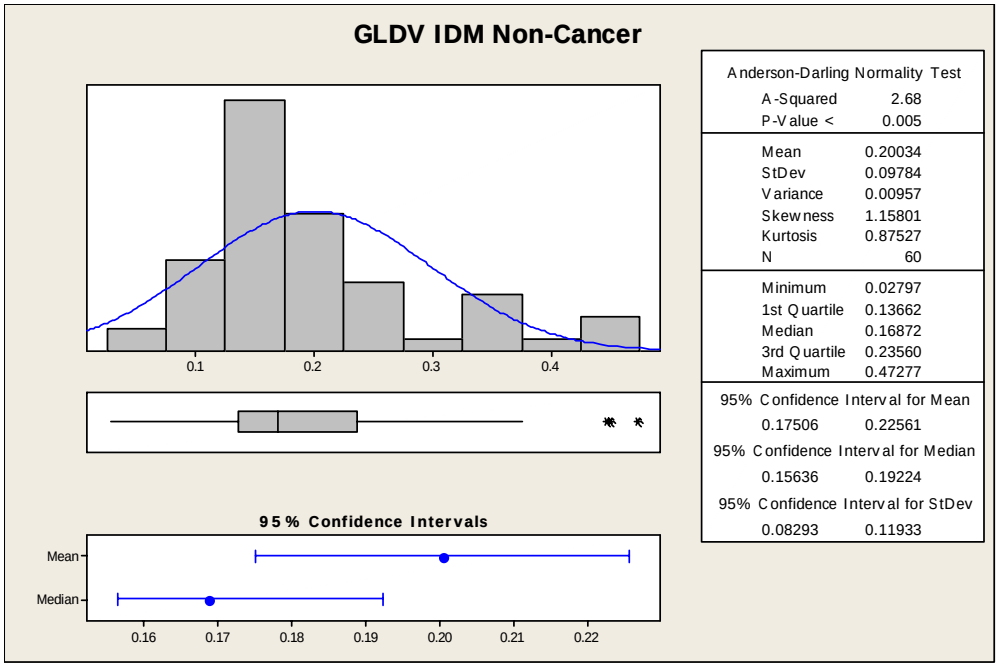


Figure 4-9 GLDV IDM Histogram and Normal Fit for Non-Cancer Regions

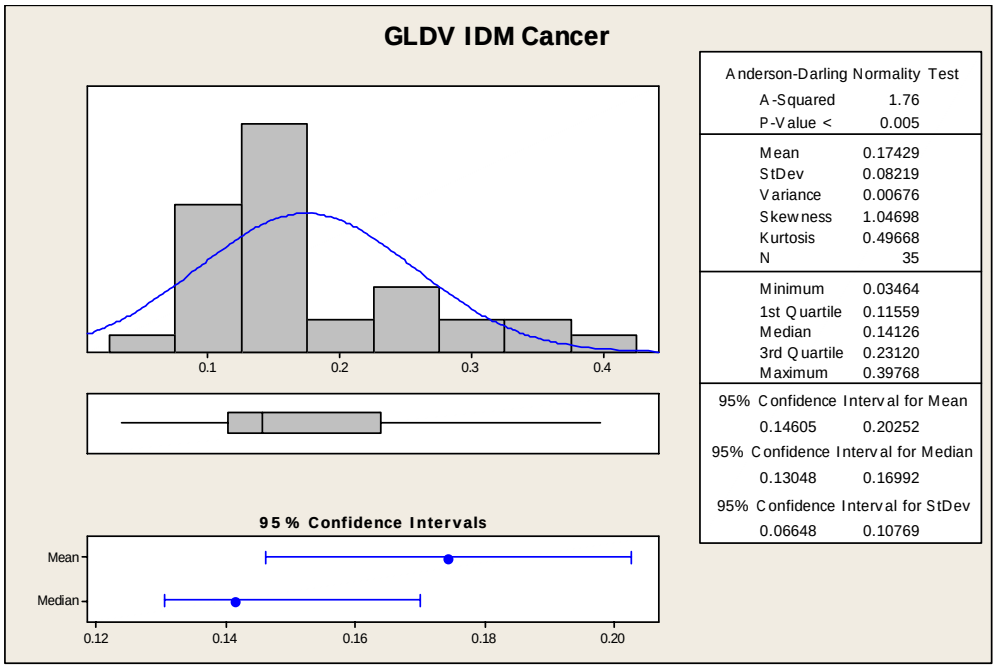


Figure 4-10 GLDV IDM Histogram and Normal Fit for Cancer Regions

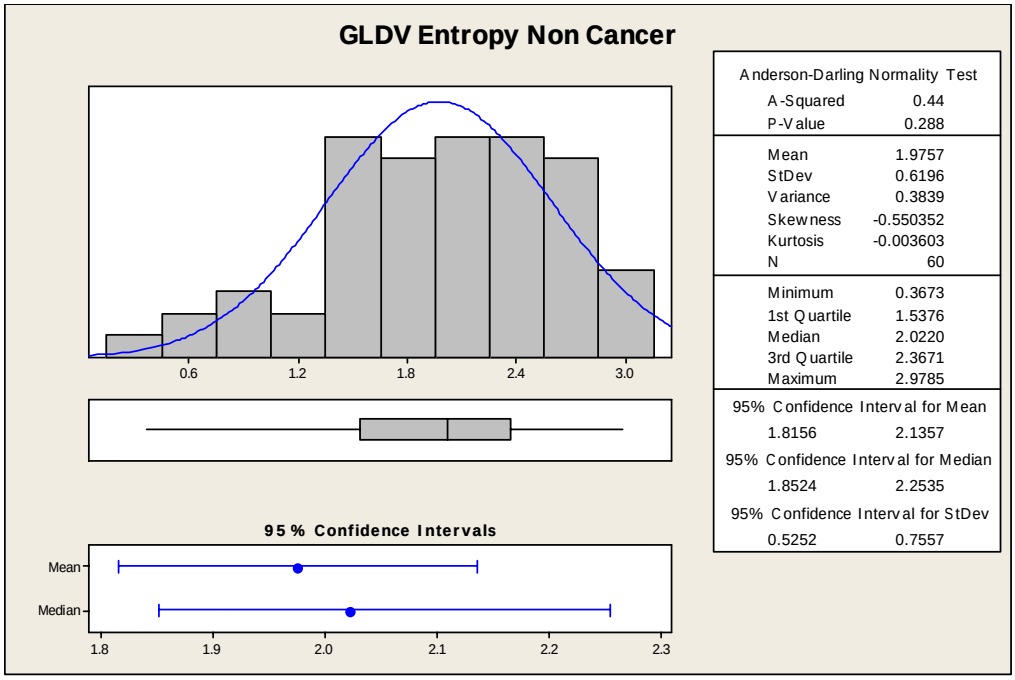


Figure 4-11 GLDV Entropy Histogram and Normal Fit for Non-Cancer Regions

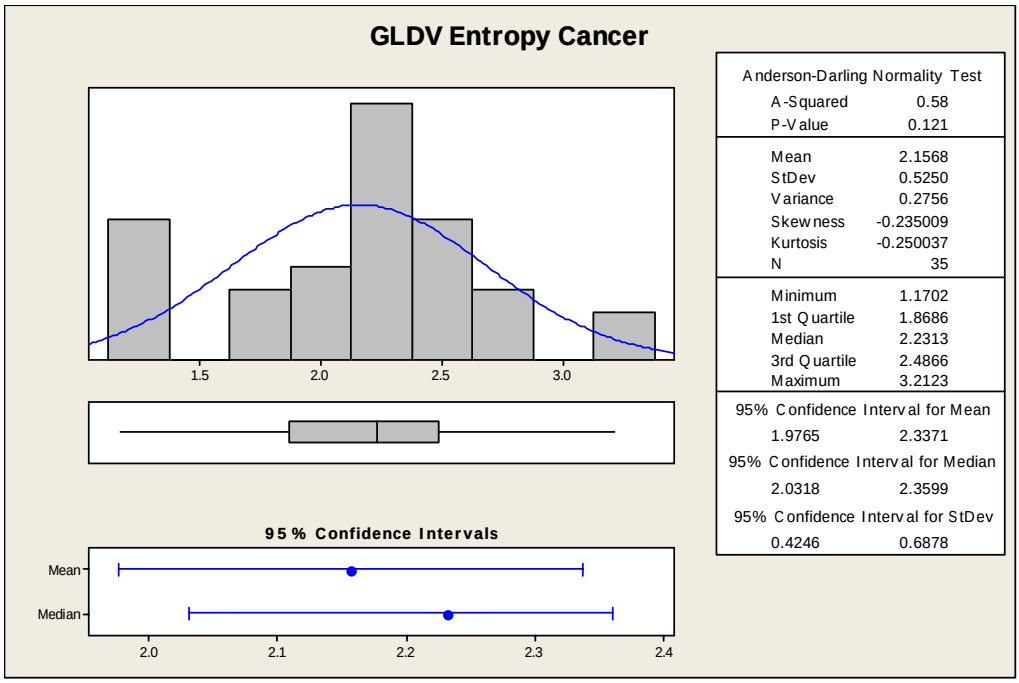


Figure 4-12 GLDV Entropy Histogram and Normal Fit for Cancer Regions

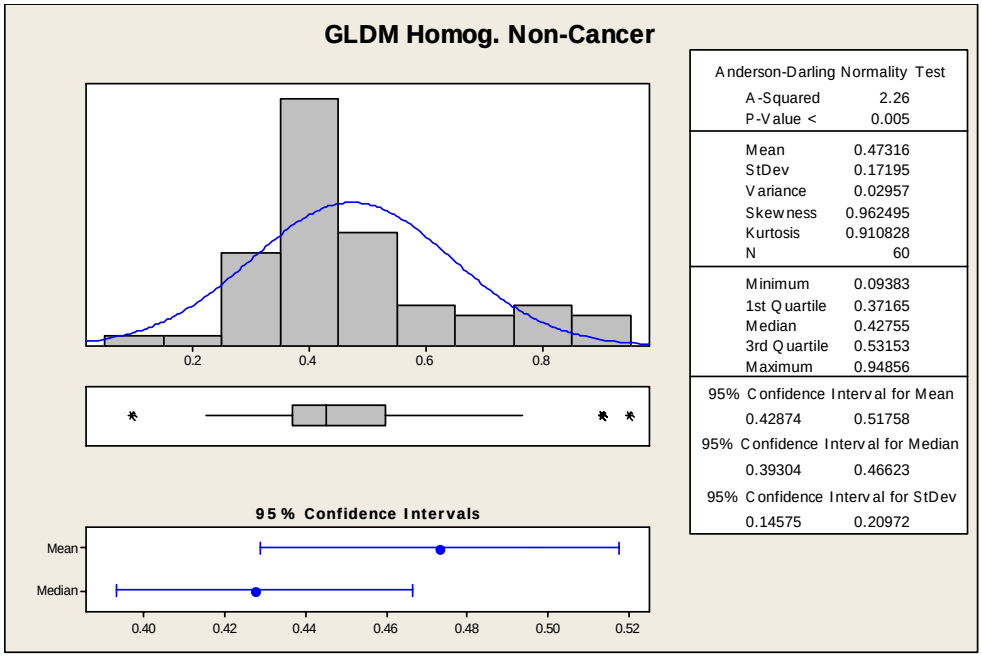


Figure 4-13 GLDM Homogeneity Histogram and Normal Fit for Non-Cancer Regions

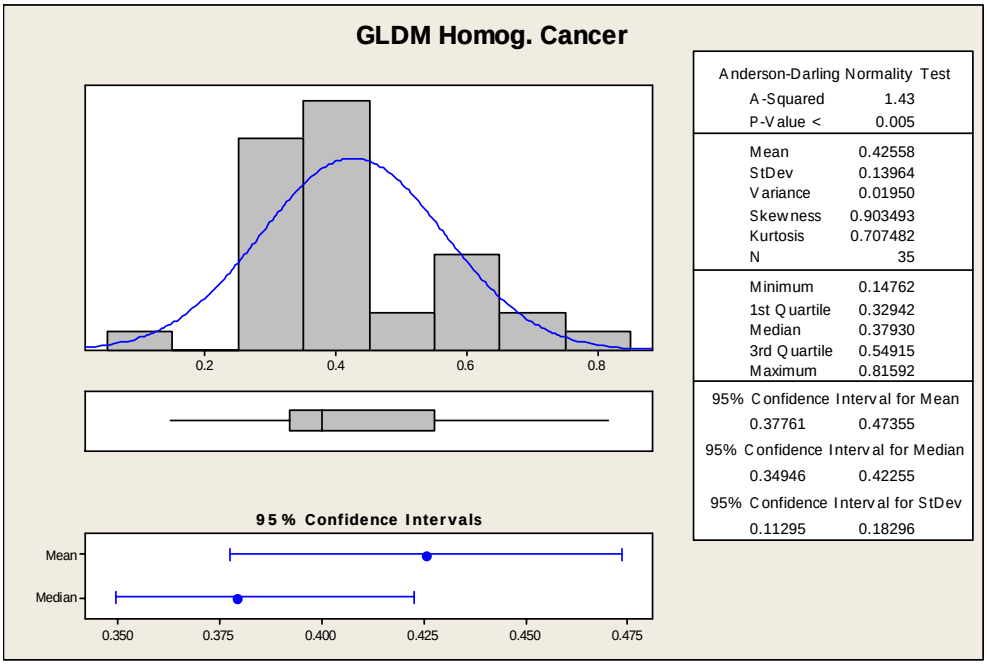


Figure 4-14 GLDM Homogeneity Histogram and Normal Fit for Cancer Regions

Furthermore, the boxplots of some features are shown in Figure 4-15 and Figure 4-16. Using the Contrast feature boxplot, the GLDV and GLDM contrasts' relation to the class (cancer = 1 or non-cancer = 0) are compared in Figure 4-15. The GLDV and GLDM Entropies' relation to classes are also shown in Figure 4-16. The line drawn across each box as mentioned earlier indicates the median, or middle, of the data. The bottom and top edges of the box mark the first (25th percentile) and third (75th percentile) quartiles, respectively. From those two figures it is clear that both the contrast features (GLDM and GLDV) range, median and distribution is quite different from the cancer regions than the non cancer regions. Conversely, the entropy features don't have this distinction between the boxes of both the cancer and non-cancer regions. Furthermore, most of the figures shown and the obtained *Kurtosis* and *A-squared* show that the Cancer regions features are closer to the normal distribution than their non-cancer counterparts are.

This statistical analysis for the constructed second order features show that those second order statistical features need to be studied more thoroughly in order to obtain the hidden information in each feature and test its discriminatory power and how this power is enhanced by combining several features together which is the task of the following chapter.

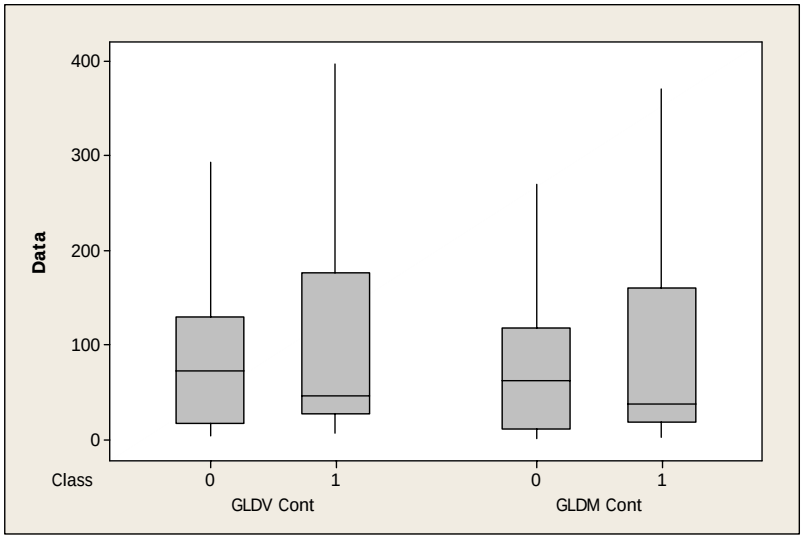


Figure 4-15 Box plot for the GLDV contrast, The GLDM contrast versus the Class

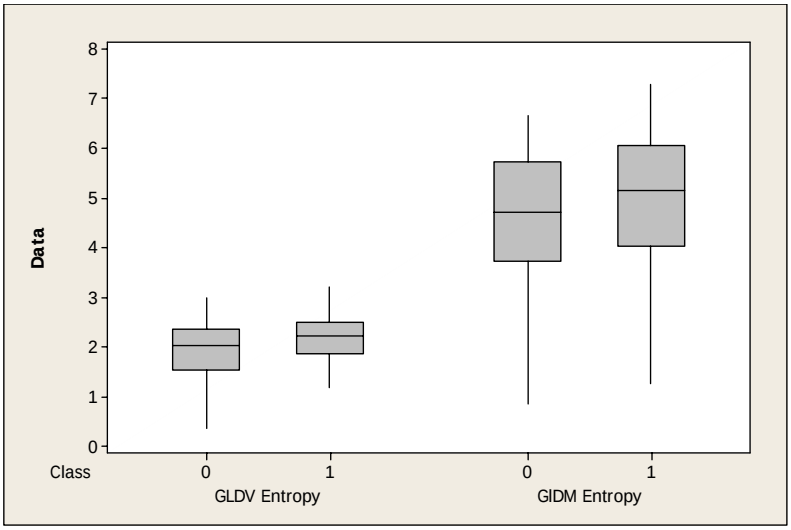


Figure 4-16 Box plot for the GLDV Entropy, the GLDM Entropy versus the Class

As mentioned above, the obtained figures and statistical measures look quite different; however, it is not an easy task to decide whether a region is cancerous or not based on visual inspection. Therefore, a need for a classifier is important to classify the regions according to these informative features.

4.6 TRUS ROIs Classification

In the previous sections in this chapter, different statistical features are constructed for each identified ROI. These features were analyzed by obtaining some statistical measures that seemed to be different from the cancer to the non-cancer regions. This observation proved that the statistical features carry information about the class, which recommends using classification techniques to benefit from this information. The different constructed feature sets were examined using three different classifiers to find out the most suitable classifier for the problem under study, these classifiers are:

- Condensed k-Nearest Neighbor (CNN) was chosen for its simplicity and fast results.
- Decision Tree (DT) was chosen for its ability to deal with data based on its information content and for visualizing the classification criteria.
- Support Vector Machines (SVMs) was chosen for its ability to deal with non-linearly separable data.

These classifiers were assessed to choose from them a classifier that is used for the rest of this thesis; those classifiers are briefed in the next sub-sections.

4.6.1 Condensed k-Nearest Neighbor (CNN)

The Condensed k-nearest neighbor classifier was applied in this work for its simplicity and quick results. The KNN rule assigns an unclassified sample to the same class as the k nearest stored

samples. The CNN retains the same approach with less time complexity. The CNN uses the following algorithm to assign the consistent subset of the original sample set. A consistent subset is a subset that when used as a stored reference set for the KNN rule correctly classifies the rest of the sample set [106]. In the CNN algorithm, a STORE bin and a GRABBAG bin are assigned and the following procedure is achieved:

1. Place first sample in STORE.

2. Classify the second sample using the KNN rule, which assigns an unclassified sample to the same class as the k nearest stored samples, using the contents in the STORE as a reference set. If the second sample is classified correctly it is placed in the GRABBAG, else it is placed in the STORE.

3. The i^{th} sample is classified similarly by the current contents of the store.

4. The final contents of the STORE are passed to the KNN rule to be used as reference points and samples in the GRABBAG are discarded.

This algorithm is evaluated in [107] and showed excellent classification results for the randomized STORE GRABBAG algorithm, which is used in this work for testing the constructed features. The results of the classification using CNN are discussed in section 4.7.

4.6.2 Decision Tree (DT)

Decision Tree is a very favorable classifier as it gives logic rules for the generated classification; moreover, it relies at each stage of the training on the information gain. Therefore, a decision tree is a recursive structure for expressing classification rules. Decision trees are systems that use a top-down, divide-and-conquer strategy that partitions the given set of objects into smaller and smaller subsets in steps with the growth of the tree [108, 109]. The decision tree classifies the data to probably cancerous and probably non-cancerous classes. It is constructed of leaf nodes and decision nodes that identify a test among one of the available features, which is called an attribute. All possible tests were examined at each stage and the test that maximizes the information gain is chosen as a step in constructing the decision tree. The algorithm can be summarized as follows and illustrated in Figure 4-17.

- Generate a decision tree from a set S of objects, each belonging to one of the classes $C_1, C_2, C_3, \cdots C_k$
- If all the objects in S belong to the same class, the decision tree of S consists of a leaf labeled with this class.
- Else, let T (feature or attribute) be some test with possible outcomes $O_1, O_2, O_3, \cdots O_k$. Each object in S has one outcome for T so the test partitions S into subsets $S_1, S_2, S_3, \cdots S_k$ where each object in S_i has outcome O_l for T, the choice of which feature to start with depends on the information gain. The feature with higher information gain is chosen as the test feature of

the node. T becomes the root of the decision tree and for each outcome O_i ;a decision tree is built by invoking the same procedure recursively on the set S_i .

The DT classifier is among the classifiers applied in this work for feature evaluation and the accuracy results are shown in section 4.7.

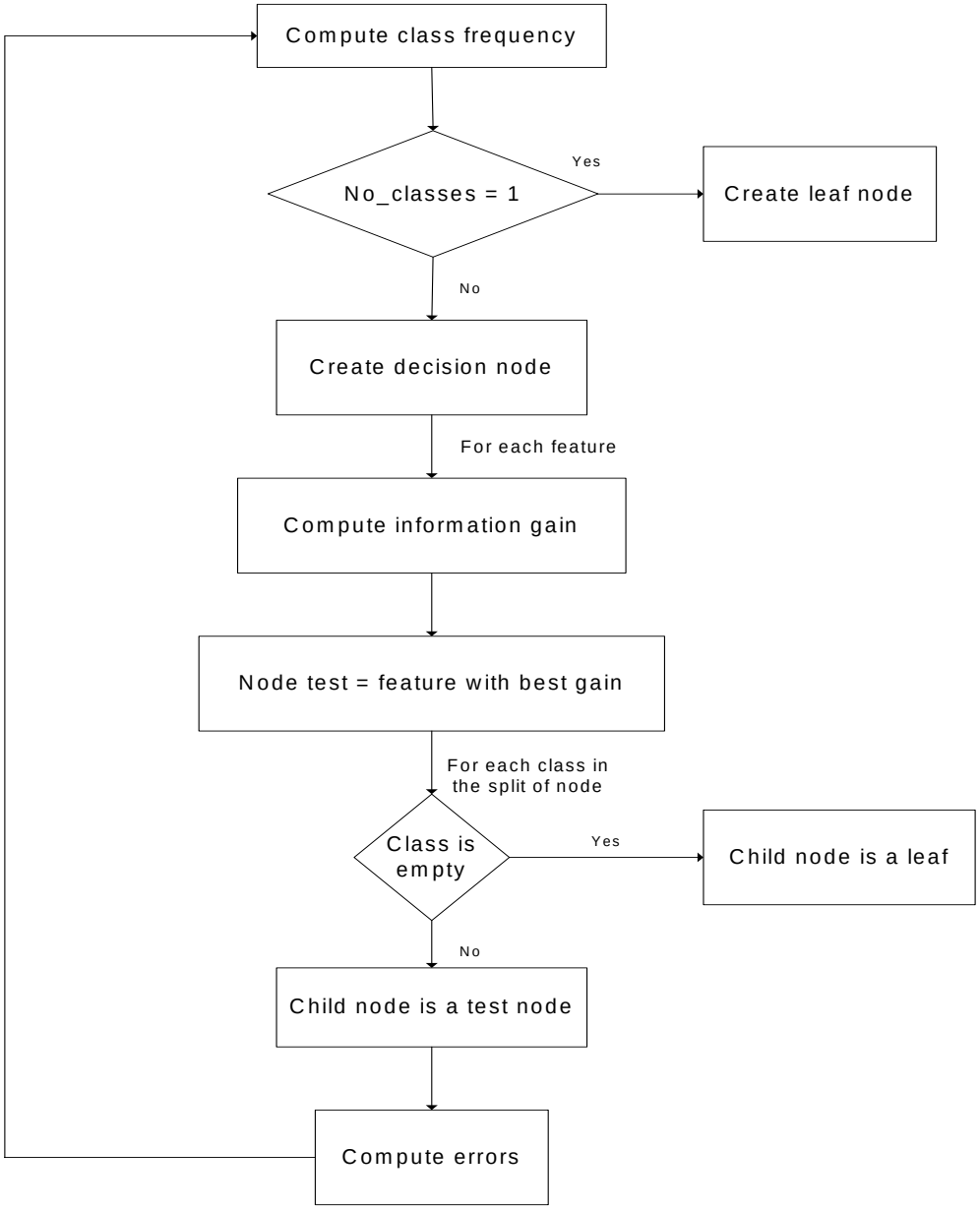


Figure 4-17 The DT Algorithm

4.6.3 Support Vector Machines (SVMs)

SVMs was first introduced by Boser and Vapnik in 1992 for the special case where data can be separated without errors and was then generalized for non-separable data sets by Vapnik in 1995 [110, 111]. Support Vector Machines were found to be an influential methodology for solving non-linear classification problems [109]. Since we are facing a non-linear problem, then applying SVMs will be appropriate for this work. The SVMs basically implements the following idea: First the input vectors are non-linearly mapped to a higher dimension feature space. Next, in this feature space a linear decision surface is constructed. Special properties of the decision surface ensure high generalization ability of the learning machine.

Therefore, SVMs depends mainly on pre-processing the data to represent patterns in a higher dimensionality space, which is achieved with a suitable non-linear mapping φ (.). In the new higher dimensional feature space, data from two classes are always separated by a hyper-plane. The hyperplane is selected so that it maximizes the margin between the two classes. It is expected that the larger the margin is the better the classification. The vectors (samples) defining the hyperplane are called Support Vectors. Therefore, the support vectors are the training samples that define the optimal separating hyper-plane and are the most difficult patterns to classify.

4.7 Classification Results and Discussion

The second order statistics is the most popular feature set for the application of tissue classification. Moreover, the second order statistics features are used for judging the classifiers, as they are the limit of the radiologist's vision (our available gold standards). Hence, the classifier that attains the best accuracy for the second order statistical features will be used for the rest of the thesis. Next, the selected classifier is used to compare the discriminatory power of different second and higher statistical feature sets and the results are shown in the following sub-sections. The cancer and non-cancer attributes in the confidence matrices represent the suspicious cancerous cases and the non-suspicious cancerous cases according to the radiologist initial assessment. Three different accuracy measures, which are Specificity, Sensitivity and overall Accuracy, were used for features as well as classifiers assessment throughout the thesis

Sensitivity: the probability that the classification result correctly identifies a cancerous ROI.

Specificity: the probability that the classification correctly indicates a non-cancerous ROI

Accuracy: the probability of any misclassified ROI.

4.7.1 Second Order Statistics Features Classification Results

The second order statistics feature sets are the ones proved earlier to possess good recognition ability in the case of ultrasound images for several applications such as fetal lung maturity [101], liver tissue classification [40, 112] and prostate tissue characterization [69, 70, 71]. Therefore, GLDM texture feature set as well as GLDV texture feature set, are used to evaluate the performance of the different

classifiers and the results are shown in this sub-section. A set of 108 regions were used in this study where 90 regions were used as a training set and the remaining 18 regions were used as the test set.

4.7.1.1 Condensed k- Nearest Neighbor (CNN)

The results of applying CNN classifier to the GLDM features, GLDV features with three neighbors are shown in Table 4-2. It is clear from the shown confidence matrices that combining all the features has better performance than using either feature set. However, the results are not as satisfactory as one expects because the performance of the CNN algorithm is limited by the distance between each incidence and in our case, these features carry information that can be captured by non-linear classifiers. This is the main reason behind choosing the DT that relies mainly on the information gain of each feature and the SVMs that can deal with non-linearly separable data.

Table 4-2 Classification results using Second order features and CNN

GLDM		Cancer	Non-cancer
	Cancer	4	2
	Non-cancer	2	10
Sensitivity = 66.67%; Specificity 83.3%; 77.7% Accuracy			
GLDV		Cancer	Non-cancer
	Cancer	4	2
	Non-cancer	2	10
Sensitivity = 66.67%; Specificity 83.3%; 77.7% Accuracy			
All		Cancer	Non-cancer
	Cancer	4	2
	Non-cancer	1	11
Sensitivity = 66.67%; Specificity 91.67% %; 83.3% Accuracy			

4.7.1.2 Decision Tree (DT)

The second order statistics feature vectors were used with the DT classifier and a sample decision tree obtained for GLDM features is shown in Figure 4-18. The classification results are shown in Table 4-3. From the table, it is clear that the GLDV features performed better Specificity than both the GLDM feature set and the combined feature set with the same Sensitivity. This makes it favourable, as it requires less computational effort. The overall accuracy of the DT classifier makes it also favourable to use GLDV. From the table, it is also obvious that the DT performed much better than the CNN classifier due to the DT's ability to classify data that are non-linearly separable. On the other hand, the nearest neighbor rule that is applicable for the case of CNN seems to be insufficient to classify the prostate texture features accurately. From Figure 4-18, the favourable feature of the DT is clear as it sets rules for its classification results. However, in our case these rules do not have a physical meaning for the doctor who does not need to go into the details of the decision making

process. However, this classifier will be very useful if the features used are the clinical data such as the age, race, family history, etc …

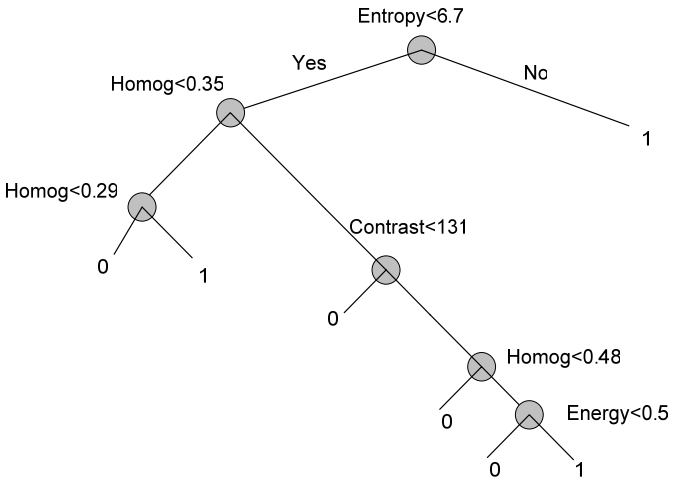


Figure 4-18 DT Sample using GLDM features

Table 4-3 Classification results using Second order features and DT

GLDM		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	2	10
83.33% Sensitivity; 83.33% Specificity; 83.33% Accuracy			
GLDV		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			
All		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	2	10
83.33% Sensitivity; 83.33% Specificity; 83.33% Accuracy			

4.7.1.3 Support Vector Machines (SVMs)

The results of applying SVMs to the available feature sets are shown in Table 4-4. From the table it is evident that the SVMs obtained almost the same classification results as those obtained using the DT

classifier for the GLDV feature set while it obtained better classification results in case of both the GLDM and the combined feature sets. Furthermore, unlike the DT, the SVMs did not generate any rules that are meaningless for the user. Therefore, SVMs classifier will be used in the rest of this thesis to compare between different features due to its superior ability to deal with non-linearly separable data.

It is clear from the classification accuracy in the case of the SVMs that using the GLDV features obtained the best results among the GLDM features and the combined features. This proves that some information that is present in the GLDM features is confusing the classifier either by being redundant or correlated. Another explanation for this result is that the GLDV features include the Mean Avg feature, which might constitute to the information needed for classification. Therefore, using a feature selection algorithm is expected to be vital in order to get rid of the redundancy and correlation between features and to keep only the informative features, which eventually will lead to the selection of the best feature subset for the tissue classification problem. Therefore, feature selection might play a significant rule in enhancing the classification accuracy. The feature selection issue is discussed and applied for the constructed statistical feature sets in the following chapter.

Table 4-4 Classification results using Second order features and SVMs

GLDM		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	1	11
83.33% Sensitivity; 91.6% Specificity; 88.8% Accuracy			
GLDV		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			
All		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	1	11
83.33% Sensitivity; 91.6% Specificity; 88.8% Accuracy			

4.7.2 Second and Higher Order Statistics Features

The statistical feature sets (GLDM features, GLDV features, GTRLM and NGTDM features) are tested by the SVMs classifier and the results are shown in Table 4-5. The same training and test sets as the previous subsections were used. It is lucid from the table that the second order statistical features obtained better classification results than the higher order statistical features. This result is rather expected in our situation because with the available gold standards used for training the classifier and testing it, at this stage, the best reached target is to mimic the expert radiologist. Therefore, we cannot outperform the radiologist's visual system that is limited to recognizing textures

with different second order statistics. Textures with the same second order statistics and different higher order statistics usually escape from the human vision discriminatory power. Being bound by the human visual system makes recognizing what is beyond the HVS hard to achieve. This shows that the higher order statistical features carry more information than what we can see and interpret with the naked eye from the TRUS image. Therefore, it is much more useful to use this kind of features in the case if the gold standard is pathology-identified regions.

Among the second order feature construction methods the GLDV features performed better than the GLDM features. This proves that as mentioned earlier, there is either extra information in the GLDM features that confused the classifier or less information as the GLDV feature set includes the Mean Average feature. Therefore, a kind of feature selection among the statistical features is important to get rid of the redundant and correlated information in order to overcome the curse of dimensionality and to integrate information from the best discriminatory feature set.

Table 4-5 Classification results using Second and Higher order features with SVMs

GLDM		cancer	Non-cancer
	cancer	5	1
	Non-cancer	1	11
83.33% Sensitivity; 91.6% Specificity; 88.8% Accuracy			
GLDV		cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			
GTRLM		cancer	Non-cancer
	cancer	4	2
	Non-cancer	1	11
Sensitivity = 66.67%; Specificity 91.6%; 83.3% Accuracy			
NGTDM		cancer	Non-cancer
	cancer	4	2
	Non-cancer	1	11
Sensitivity = 66.67%; Specificity 91.6%; 83.3% Accuracy			

4.8 Conclusions

In this Chapter, different statistical textural feature sets were constructed. These feature sets contain the well-established second order statistical features as well as the higher order statistical features that were not applied for ultrasound images before. Different classifiers were also compared to select a classifier that is suitable for the available feature sets and will be used for the feature analysis task throughout the rest of the thesis. The SVMs was selected due to its ability to deal with this kind of non-linearly separable data, which led to excellent classification accuracy. Using the second order statistical features obtained better results than their higher order counterparts did because we are bound at this stage by the radiologist's vision that is limited only to second order statistics. Therefore, second order statistical features are recommended at this stage. However, a framework was developed in this chapter using higher order statistics features to obtain information from the TRUS images in the case where pathology identified data is available. Moreover, several statistical measures were calculated for the constructed second order statistical features among them are each feature's distribution and how far is it from the normal distribution. From the statistical measures obtained, it was concluded that the statistical features carry information about the tissue class; therefore, feature selection is a crucial step to highlight the most powerful features that carry more information about the region's tissue characteristics. Feature selection will be discussed in the following Chapter.

Chapter 5
Dimensionality Reduction

5.1 Introduction

In the previous Chapter, different statistical features were constructed from the obtained ROIs. These features were analyzed statistically and it was observed that there is an obvious difference between the statistical measures of the cancer and non cancer regions' histograms. Moreover, three different classifiers were evaluated using the constructed second order statistical features. The SVMs classifier was chosen in this work due to its ability to deal with non-linearly separable data. Moreover, the second and higher order statistical features were examined using the SVMs where the second order features obtained better classification results for the available data set. Furthermore, it was also found from the statistical measures and histograms that each feature is carrying certain amount of information that can reflect the tissue type. The obtained statistical measures highlighted the fact that there is a difference between cancerous and non-cancerous ROIs' features. Yet, this difference in the statistical measures is not enough to conclude which features have more discriminatory power of the tissue texture. However, the integrated use of these information contents is expected to lead to a better tissue typing. Therefore, finding hidden patterns in this data has become an important issue of the CAD system proposed in this thesis. The constructed features statistical analysis described in the previous chapter proved that the feature selection process is crucial to obtain the best feature subset for the purpose of tissue classification. It is well known that pre-processing data, before using any learning or discovering algorithm, is an essential step in the success of this learning algorithm [117, 118]. Moreover, the feature sets constructed in the previous chapter obtained different classification results, which might be enhanced by pre-processing these feature sets. Therefore, it is useful to take the best subset out of these features in order to obtain accurate ROI diagnosis. This dimensionality reduction principle is investigated in this chapter.

Generally, feature analysis is an important issue in reducing the available feature space that is represented by pixel values for the purpose of CAD in medical images. Feature Construction (FC), Feature Extraction (FE) and Feature Selection (FS) are effective approaches to data reduction whose goals are reducing the size of the feature set, focusing on its relevant components, and improving its quality. Feature construction creates more features and discovers missing information from the relationships between features. FC was applied in the previous chapter to construct statistical features from the identified ROIs and will be used in the coming chapters to construct some other texture features to aid in cancer recognition using the TRUS images. It is noted in [119] that many algorithms have poor results in the case of large number of irrelevant or redundant features. Pre-processing, which include FE and FS before classification is a good option to deal with huge data sets. Different feature analysis avenues are shown in Figure 5-1.

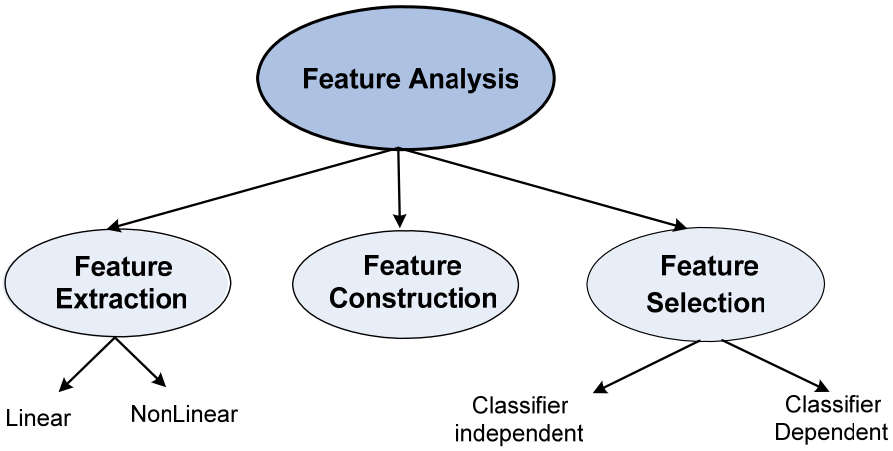


Figure 5-1 Feature Analysis Avenues

Figure 5-2 and Figure 5-3 illustrate a simple example of how the features' quality and redundancy can affect classification results. It is clear from Figure 5-2 that the classes w_1 and w_2 are completely separable under feature 1; therefore, feature 2 is considered redundant because it didn't contribute any discriminatory information to the feature set. While in Figure 5-3 the center clusters are uncertain under both features but each feature adds some separability where feature 2 provides more separability than feature 1.

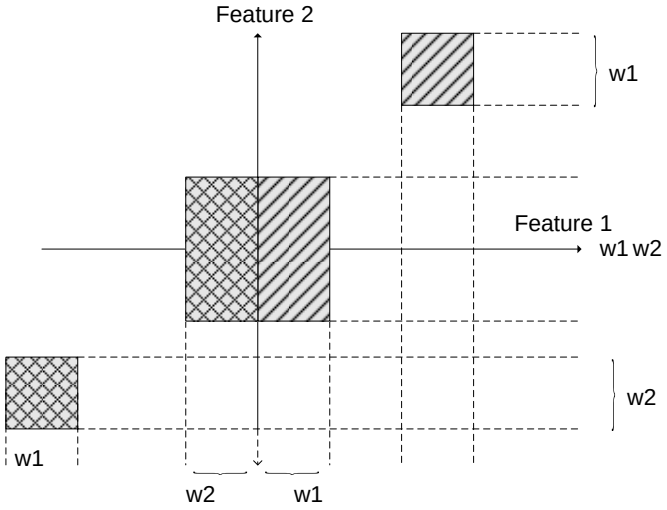


Figure 5-2 Features' Redundancy

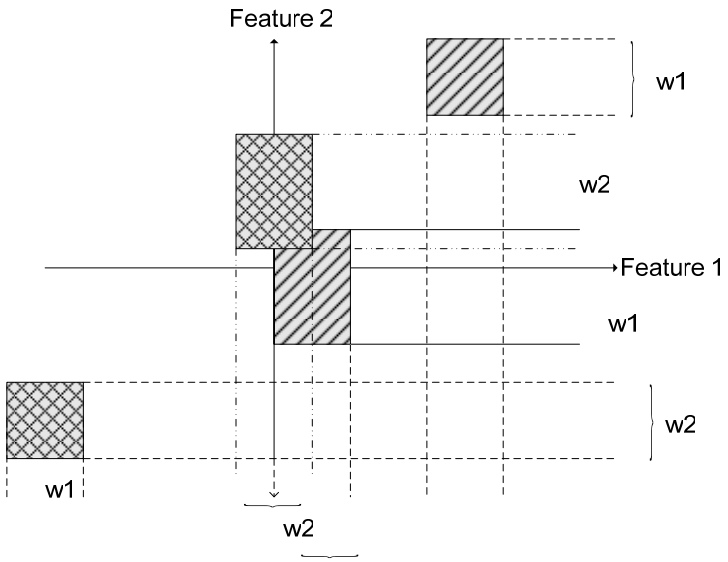


Figure 5-3 Features' Quality

It should also be noticed that the available data set for this work is limited (only 33 TRUS images), which might lead to an overfitting problem where the number of available observations is inadequate to the number of available features. The solution to this problem is either selecting an appropriate subset from the existing features (Feature Selection) or combining some of the available features in some way to obtain a new feature set (Feature Extraction) [109]. Therefore, dimensionality reduction is an important stage in this thesis to achieve accurate tissue classification.

In the case of TRUS image feature analysis the correlation measure is calculated among some of the obtained features in order determine how correlated these features are. The correlation among features is shown and explained in subsection 5.3.1.

5.2 Feature Extraction FE

FE means the generation of new set of features from the original ones using some transformation function. A common example of linear transformation FE is the Principal Component Analysis (PCA) in which high dimensional data is projected into a lower dimensional space using the Eigen decomposition [109].

FE was usually applied for dimensionality reduction for either further processing [118] or visualization [119]. FE is utilized and explained in this work in Chapter 7 where Locally Linear Embedding systems (LLE) are used to extract new salient features from the constructed ROIs texture feature sets.

5.3 Feature Selection

The output of the previous chapter is feature sets that contain many ROIs statistical features from all the previously explained statistical feature construction methods. These feature sets were applied to the SVMs classifier that identified some classes (e.g., cancerous and non-cancerous). These feature sets might have some redundant and correlated features (curse of dimensionality) which is the main motive for using the feature selection techniques in order to enhance the features' quality.

Large number of features (more information) tends to increase classification errors, and the classifier structure becomes more difficult to interpret. This was clear in the classification results using the GLDM with the GLDV features compared to the classification results using GLDV features alone which was explained in section 4.7 . Further, the classifier is sensitive to outliers and might face an overfitting problem that leads to less classification accuracy. Moreover, there is no guarantee that the classifier will perform as well on a new set of samples.

The principle of feature selection is to take a set of candidate features and select a feature subset, which retains the information needed for pattern classification. In some cases, it is possible to derive a subset of features, which forfeit none of the information needed for classification, yet has minimal correlation and common information among the selected features. This process usually leads to better classification accuracy and the subset of features is referred to as an *optimal subset.* The optimal subset sometimes results in reducing the minimum probability of error, when a decision rule is applied in both the original and the subset space.

FS is used to select a subset of features from a given set of features, which leads to performance enhancement of the recognition system. In the worst case, exhaustive search is applied to guarantee the global optimal feature subset. The exhaustive search usually examines all possible subsets of features, which grows exponentially with the number of original features. Therefore, its' time complexity is exponential on the dimension of the feature space, which causes this approach to be impractical even for medium number of features. One of the most traditional algorithms that is less expensive than the exhaustive search is the Branch and Bound (BB), which is an optimal method for monotonic feature sets (where the accuracy increases with the increase of number of features). Nevertheless, its computational cost is still prohibitive for large feature spaces. Moreover, the different feature sets that are generated in our case are not considered monotonic (adding new feature doesn't necessarily improve the classification result) [120], which makes the BB algorithm not adequate for the case of TRUS image texture features. Feature selection is categorized into two main categories, namely, classifier dependent and classifier independent feature selection techniques. In the recent years, many feature selection algorithms have been developed, but no optimal algorithms can actually be suitable for all problems [121]. Therefore, in this research both avenues are applied and the one that better fits the TRUS image texture features application is ideal to be used for the rest of the thesis.

5.3.1 Texture Features Correlation

In the previous chapter, some statistical measures are calculated for the constructed TRUS image features. Those calculated measures reflect the information each feature carry about the ROI tissue class. However, these features might be correlated or might carry the same information, which causes classification errors. Therefore, in this section the correlation coefficients among some of those features are calculated. In probability theory and statistics, *correlation*, also called correlation coefficient, indicates the strength and direction of a linear relationship between two random variables. Generally, correlation refers to the departure of two variables from being independent [113, 114].

Correlation among the constructed second order statistical features that were explained in section 4.3.1 is calculated and the correlation coefficients are shown in Table 5-1. The correlation coefficient measures the degree of linear relationship between two variables. The correlation coefficient assumes a value between -1 and +1. If one variable tends to increase as the other decreases, the correlation coefficient is negative. Conversely, if the two variables tend to increase together the correlation coefficient is positive. However, when performing a correlation analysis it should be clear that correlation coefficients only measure linear relationships. A meaningful nonlinear relationship can exist even if the correlation coefficient is zero [114].

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \qquad \textbf{5-1}$$

where:

$\bar{x}$ = sample mean for the first variable

s_x = standard deviation for the first variable

$\bar{y}$ = sample mean for the second variable

s_y = standard deviation for the second variable

n = column length

From Table 5-1 it is clear that there is some high correlation coefficients between the ROIs' second order texture features, which recommend the use of FS techniques to choose a subset of uncorrelated features that, will result in enhancing the classification accuracy.

A plot of two highly positive correlated features (GLDV Contrast and GLDV Mean Average) is shown in Figure 5-4 where both variables increase concurrently, therefore, a positive linear relationship exists. The points on this plot hug the line closely, suggesting that the relationship between the variables is strong. The correlation coefficient for this relationship is +0.943. A plot of two highly negative correlated features (GLDV ASM and GLDM Entropy) is shown in Figure 5-5, where the points on this plot hug the line closely (but less closer than the previous figure), suggesting that the relationship between the variables is strong. The correlation coefficient for this relationship is

-0.837 which means when any of these features increase the other feature usually decreases. Moreover, a plot for two almost uncorrelated features (GLDV Contrast and GLDM Homogeneity) is shown in Figure 5-6 where the data points in this plot appear to be randomly distributed. They do not fall close to the line indicating a very weak relationship if one exists. The correlation coefficient for this relationship is -0.046.

Therefore, from the previous discussion it is clear that there is a great need to get rid of the redundant features and eliminate the existing correlation. However, the correlation coefficients cannot be considered the only measure for interdependence among features as it focuses only on the linear relationships [114, 116]. Therefore, feature selection algorithms should be studied and applied to achieve this task which is the main topic discussed in the coming sections.

Table 5-1 Correlation among Second Order Statistical Features

		GLDV					GLDM		
		Contrast	Mean	Entropy	IDM	ASM	Entropy	Energy	Contrast
GLDV	Mean	**0.943**							
	Entropy	0.097	0.29						
	IDM	0.06	-0.18	**-0.77**					
	ASM	0.09	**-0.87**	**-0.87**	**0.93**				
GLDM	Entropy	-0.45	**0.7**	**0.7**	**-0.78**	**-0.84**			
	Energy	0.3	**-0.68**	**-0.7**	**0.86**	**0.93**	**-0.86**		
	Contrast	**1**	0.08	0.08	0.07	0.1	-0.45	0.3	
	Homog	-0.05	**-0.79**	**-0.79**	**0.99**	**0.91**	**-0.72**	**0.81**	-0.03

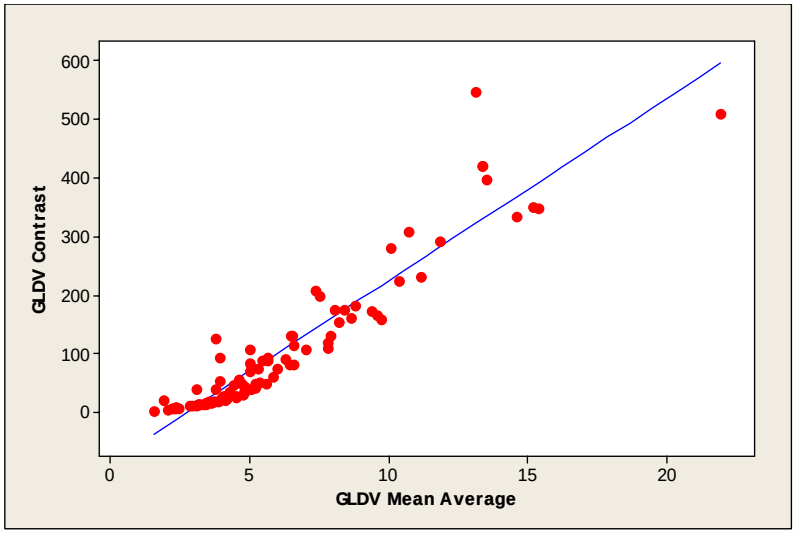


Figure 5-4 Correlation between GLDV Mean Average and GLDV Contrast

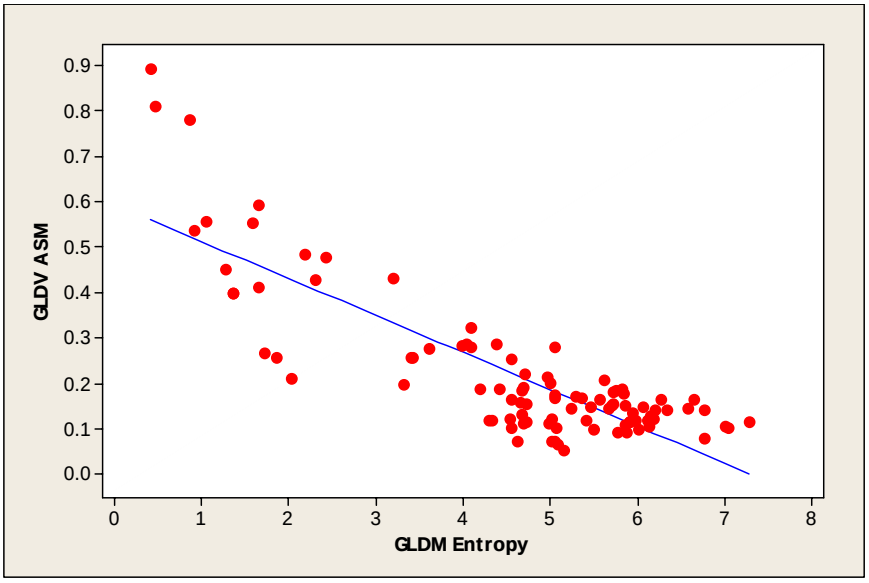


Figure 5-5 Correlation between GLDV ASM and GLDM Entropy

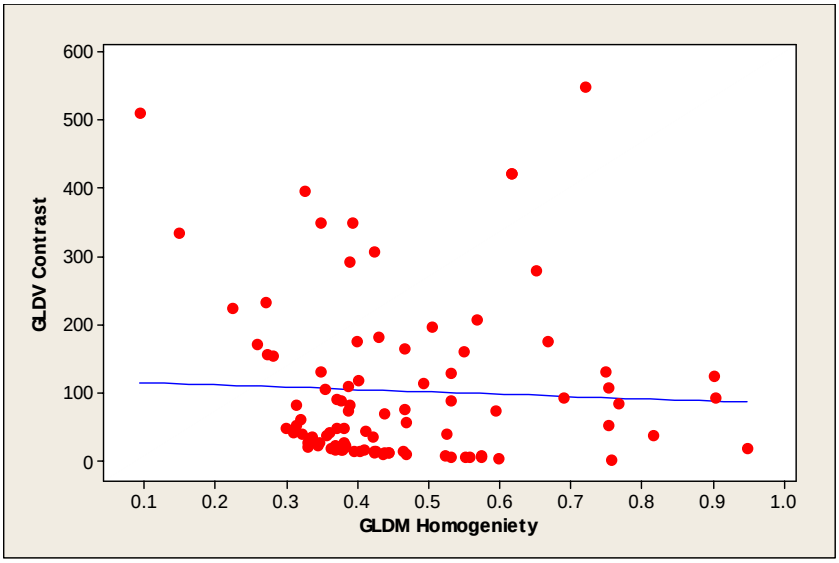


Figure 5-6 Correlation between GLDV Contrast and GLDM Homogeneity

5.4 Classifier Independent Feature Selection

Feature analysis for classification is based on the discriminatory power of features. This is a measure of the usefulness of the feature(s) in determining the class of an object. Traditional feature analysis for classification addresses only classifier based discriminatory power, i.e., a classifier is selected, and then the discriminatory power of a feature is proportional to the accuracy of the classifier when it uses that feature. Traditional feature analysis for classification is thus classifier-driven. Therefore, in order not to rely on a certain classifier, there is a need to integrate the classifier independent feature selection algorithm in the proposed framework.

As discussed in section 5.3.1, it was shown that there is a kind of correlation among the constructed features. However, this correlation is not enough for feature selection as it measures only the linear correlation. Therefore using a more generic method to determine the dependency among the features should be used to get rid of the redundant and correlated features. Moreover, the selected features should also satisfy the condition that it represents the tissue class effectively. Therefore, the information the feature carries about the class should be included in the analysis. The adequate solution to this problem is the Mutual Information Feature Selection algorithm (MIFS) which is based on the concept of Mutual Information (MI) [122, 123].

5.4.1 Mutual Information Feature Selection (MIFS)

MI can be used to evaluate the information content of each individual feature about the output class, and about each one of the other features. MI is based on probability distributions and measures the amount of information one random variable contains about the other [122, 124]. Denoting X as a

random variable describing a feature and C as a random variable describing the class, the mutual information $I(C;X)$ is a measure of the amount of information that random variable X contains about the random variable C. It can also be interpreted as the reduction of the uncertainty of one random variable due to the knowledge of the other. Thus, it provides a criterion for measuring the effectiveness of a feature for class separability. Interdependence between a feature and the classes is proportional to the value of $I(C;X)$. The same procedure can be done to determine interdependence among features; where $I(X_1;X_2)$ represent the interdependence among features X_1 *and* X_2 that should be minimized to avoid selecting two or more similar features. Therefore, we are looking for maximizing $I(C;X)$, and minimizing $I(X_1;X_2)$.

5.4.1.1 Entropy and Mutual Information

In the classifier independent feature selection problem, it is required to find inputs (features) that carry as much information about the output (classes) as possible, therefore, tools to measure information is a must. Using Shannon's information theory, there is a way to measure information of random variables using entropy and mutual information. Entropy is a measure of uncertainty of random variables. If a discrete variable X has a probability density function (*pdf*) $p(x)$ then the entropy of X is defined as:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x), \qquad \textbf{5-2}$$

While for two discrete random variables X and Y whose joint *pdf* is $p(x,y)$, the joint entropy $H(X,Y)$ is defined as:

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y), \qquad \textbf{5-3}$$

Usually when some variables are known and others are not, the remaining uncertainty is measured by the conditional entropy:

$$H(Y/X) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(y/x) \qquad \textbf{5-4}$$

The relation between the conditional entropy and the joint entropy is given by the chain rule which implies that the total entropy of random variables X and Y is the entropy of X plus the remaining entropy of Y given X. This can be expressed by equation 5-5 and can be shown in Figure 5-7:

$$\begin{aligned} H(X,Y) &= H(X) + H(Y/X) \\ &= H(Y) + H(X/Y) \end{aligned} \qquad \textbf{5-5}$$

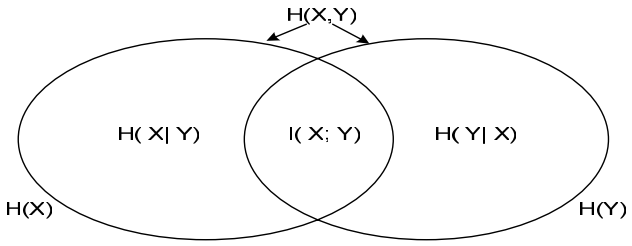


Figure 5-7 Relation between Mutual Information and Entropy

5.4.1.2 MI between Features and Classes (MIFC)

From the previous discussion, the MI between feature values and classes can be calculated as follows [122]:

$$I(C;X) = H(C) - H(C|X), \quad \textbf{5-6}$$

The higher $H(C|X)$ is, the higher the uncertainty. The MI $I(C;X)$, is maximum when the class is totally dependant on the feature, while it is minimum when the class and the feature are totally independent.

The entropy $H(C)$ depends mainly on the classes and it represents the upper limit for $I(C;X)$.

$$H(C) = -\sum_{c \in C} p(c) \log_2 p(c), \quad \textbf{5-7}$$

Where $p(c)$ is the probability density function of c.

The conditional entropy $H(C|X)$ measures the degree of uncertainty entailed by the set of classes C given the set of feature values X, and can be computed as:

$$H(C|X) = \sum_{c \in C} \int p(x,c) \log_2 \frac{p(x|c)\,p(c)}{p(x)} dx. \quad \textbf{5-8}$$

This formula is valid for the case of continuous functions. However, in the case of feature selection, the feature space is quantized into intervals of width Δx and the integration is replaced by summation as follows:

$$H(C|X) = \sum_{c\in C}\sum_{x\in X} p(x,c)\log_2 \frac{p(x|c)\,p(c)}{p(x)}\Delta x \qquad but \quad p(x,c) = p(x/c)\,p(c)$$
$$= \sum_{c\in C} p(c)\sum_{x\in X} p(x\,|\,c)\Delta x \log_2 \frac{p(x|c)\,p(c)}{p(x)}. \qquad \textbf{5-9}$$

Since the classes for which the data set belongs are known (e.g., cancerous or non-cancerous), then $p(c)$ can be calculated by dividing the number of each class by the total number of classes. $p(x)$ and $p(x|c)$ are estimated using the Parzen Density Estimator [123, 125].

5.4.1.3 MI among Features

The number of selected features should be minimized to get rid of redundancies and increase the efficiency of the classifier. In order to achieve this task, no features carrying the same information (their mutual information is relatively big) should be selected. Therefore, the MI between different features e.g. (X_1, X_2) is $I(X_1; X_2)$ should be calculated:

$$I(X_1; X_2) = H(X_2) - H(X_2/X_1)$$
$$= -\sum_{x_2\in X_2} p(x_2)\Delta x \log_2\, p(x_2)\Delta x - \sum_{x_2\in X_2} p(x_2)\Delta x \sum_{x_1\in X_1} p(x_1|x_2)\Delta x \log_2 \frac{p(x_1|x_2)\,p(x_2)}{p(x_1)}. \qquad \textbf{5-10}$$

The *pdfs* used here are also estimated using the Parzen density estimation.

5.4.1.4 Mutual Information Feature Selection (MIFS) Algorithm

MIFS was proposed in [122, 123] and the algorithm is shown in Figure 5-8. In this work, the algorithm is applied for the TRUS image texture features constructed in the previous chapter and the algorithm is summarized as:

1. *Initialization* Set X ← "initial set of N features;" S ← "empty set."
2. *Computation of the MI with the output class* for each feature $X_i \in X$ compute $I(C; X_i)$
3. *Choice of the first feature* find the feature f that maximizes $I(C; X_i)$;set $X \leftarrow X \setminus \{X_i\}$; set $S \leftarrow \{X_i\}$
4. *Greedy selection* repeat until $S = k$ (the desired number of selected features):
    a. *Computation of the MI between variables* for all couples of variables $(X_i; X_s)$ with $X_i \in X$ and $X_s \in S$ compute $I(X_i; X_s)$.
    b. *Selection of the next feature* choose feature X_i as the one that maximizes

$$I(C;X_i) - \beta \sum_{s \in S} I(X_i;X_s) \qquad \textbf{5-11}$$

c. Set $X \leftarrow X \setminus \{X_i\}$ and $X_s \leftarrow X_s \cup X_i$

5. Output the set X_s containing the selected features.

The parameter β controls the relative importance of the MI between the candidate features and the already selected features with respect to MI with the class. If β is zero, only the MI with the output class is considered for each feature selection. If β increases, this measure is discounted by a quantity proportional to the total MI with respect to the already-selected features. [122]. Different sizes of the selected feature subsets were investigated in this work. It was found out that a feature subset with a minimum size of four features results in excellent accuracy; therefore, this subset size will be used in this study. The MIFS algorithm is used in this work to select four features from the available combined set of features. Moreover, in this work β value was set to 0.5 in order to give some weight for both the MIFC and MIXY.

5.4.2 Feature Selection Results using MIFS

The MIFS algorithm explained in the previous section is used to rank the available second order statistical features according to the information content of each feature. The resulting MI between each feature and the class is shown in Figure 5-9 where the x-axis represents the feature number and the y-axis represents the mutual information between that feature and the class, while the MI among features is shown in Table 5-2.

In this work, it was found that the feature subset that maximizes MIFC and minimizes MIXY is constructed of four features. This selected feature subset performed better than the original feature sets and resulted in more accurate classification compared to the case of using all the constructed second order features combined. As mentioned earlier the available data set consists of 33 images with 108 ROIs identified. The gold standards that we use are the radiologist's marking on the TRUS images with an accuracy ceiling of 66 %. The classification results are shown in Table 5-3 where the SVMs classifier is further compared to both the CNN and the DT. The classification results obtained using the SVMs are better than the ones obtained by either the CNN or the DT, which supports the recommendation of the previous chapter to carry on with the SVMs for the rest of the classification in this work. Moreover, the selected feature subset obtained better results using the SVMs than both the GLDM feature set and all the features combined. However, it still performed as good as the GLDV features. Yet, the selected feature subset is expected to perform better than the GLDV in the case where the data set is larger and when the pathology-identified data set is available.

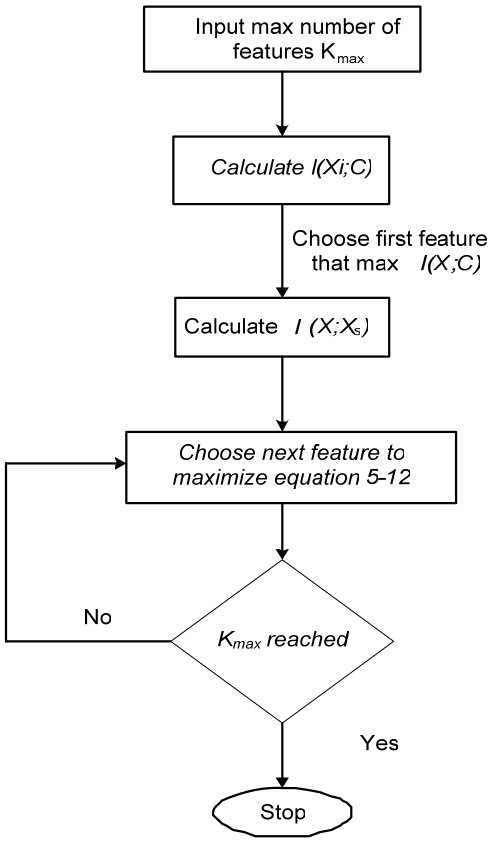


Figure 5-8 MIFS Algorithm

Table 5-2 Interdependence among features

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Feature 9
Feature 1	0	0.68	0	0	0	0	0	2.94	0
Feature 2		0	0.28	0.38	0.07	0.11	0.03	0.23	0.1
Feature 3			0	0.54	0.8	0.42	0.43	0	0.58
Feature 4				0	0.97	0.5	0.65	0	2.22
Feature 5					0	0.6	0.94	0	0.92
Feature 6						0	0.67	0	0.46
Feature 7							0	0	0.55
Feature 8								0	0
Feature 9									0

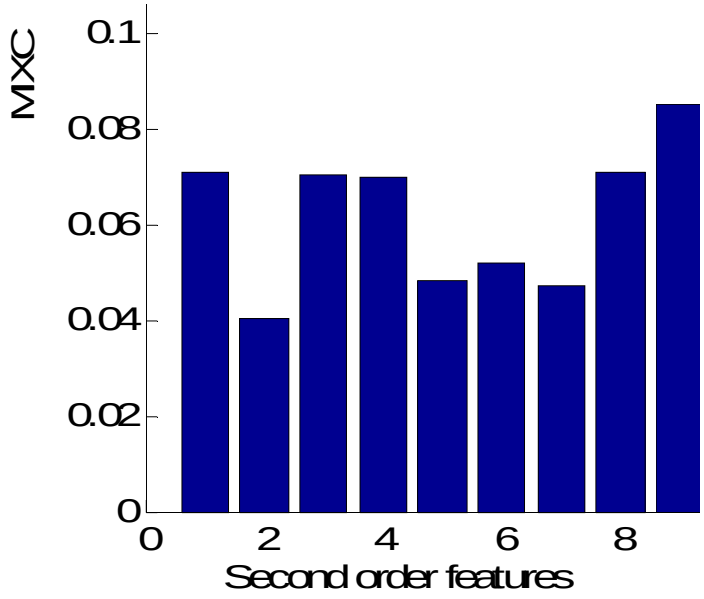


Figure 5-9 Interdependence between features and classes

Using the MIFS algorithm, we obtained excellent accuracies especially when we combined it with the SVMs classifier and the second order statistical features. However, this algorithm is very much time consuming for this limited number of features (9 features). Yet, its ability to choose the best feature subset for the problem is still acknowledged. Therefore, an idea of using other feature selection methods arises. Having chosen the SVMs as the classifier that is going to be used for the rest of the thesis, using classifier dependent feature selection methods is a promising choice. Especially that the SVMs performed as good even for the selected feature sub-set, and neither the CNN nor the DT have the same performance using the selected feature sub-set.

Table 5-3 Classification results using MI feature subset

CNN		Cancer	Non-Cancer
	Cancer	4	2
	Non-Cancer	2	10
Sensitivity = 66.67%; Specificity 83.3%; 77.7% Accuracy			
DT		Cancer	Non-Cancer
	Cancer	5	1
	Non-Cancer	1	11
83.33% Sensitivity; 91.6% Specificity; 88.8% Accuracy			
SVMs		Cancer	Non-Cancer
	Cancer	5	1
	Non-Cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			

5.5 Classifier Dependent Feature Selection

In the classifier dependent feature selection avenue, a classifier is pre-defined (SVMs in this work) and the accuracy of the classification using the selected feature subset is judged. In supervised learning, feature selection is often viewed as a search problem in a space of features. On one hand, filter methods (classifier independent) use an evaluation function that relies exclusively on data properties, thus is independent on any particular classification algorithm. This method has been explained and used in the previous section 5.4. On the other hand, wrappers (classifier dependent) use the inductive algorithm to estimate the accuracy of a given subset [126]. The wrapper methods usually choose a feature subset that satisfies a certain criterion. The difference between the features chosen by the wrapper method and those left out is characterized by a notion of relevance. In this work, the relevant feature subset is the subset that maximizes the classifier accuracy. The most common wrapper method is the Sequential Forward Search Algorithm (SFS) and the Sequential Backward Search algorithm (SBS). One of the obvious disadvantages of these approaches is that they may suffer from *nesting*. That is, the already selected features determine the course of the remaining selection process, as once an attribute is included in or removed from the feature subset, it cannot be reconsidered. Another disadvantage could be named as *feature missing*. *Feature missing* results in missing a feature if this feature alone provides little discrimination, but together with another feature they are very effective, a case that is present in the TRUS images features. The sequential algorithms may never detect this combination since either feature alone would never be selected [120, 126].

Classical Wrapper methods are widely acknowledged as a superior way in supervised learning problems. However, the disadvantage discussed earlier didn't recommend the use of these methods in the TRUS images feature selection problem. Therefore, intelligent wrapper methods are needed in order to achieve the best possible selection considering the available small data set. Two different artificial life feature selection methods are introduced in this thesis. These methods are originally used

for solving optimization problems and are tailored and adopted to fit the ROIs texture feature selection problem.

5.6 Artificial Life Feature Selection

Artificial life (Alife) is a field of study dedicated to understanding life by attempting to derive general theories underlying biological phenomena, and recreating these dynamics in another environment making them accessible to new applications. This research links biology and computer science.

Two different Alife techniques which are Particle Swarm Optimization (PSO) and Ants Colony Optimization (ACO) are tailored and adopted to be applied in this work for feature selection using the constructed statistical texture features for the identified ROIs (GLDM, GTDV, NGTDM and GTRLM). These methods are compared and evaluated based on the classifier accuracy.

5.6.1 Particle Swarm Optimization (PSO)

Particle swarm optimization is inspired from artificial life (Alife) in general such as bird flocking, fish schooling, and swarming theory in particular [127]. PSO comprises a very simple concept, and it requires only primitive mathematical operators, which makes it computationally inexpensive in terms of both memory requirements and speed.

PSO is a population-based search algorithm that is initialized with a population of random solutions, called particles. Each particle in PSO moves through the search space with a velocity that is dynamically adjusted according to its own and its companions' historical behaviors [128]. The particles have a tendency to move toward better search areas over the course of a search process. Acceleration or velocity is weighted by a random term, with separate random numbers being generated for acceleration toward the best solution obtained by that particle (*pbest*) and the best group solution obtained any other particle (*gbest*). PSO was introduced in 1995 [127] and since then it has attracted a lot of attention from researchers around the world. The focus of the work done in this thesis is on the application of PSO in the feature selection of the TRUS image ROIs texture features.

5.6.1.1 PSO Algorithm

Originally, PSO was designed for continuous functions optimization problems. The algorithm has now been extended to cover binary and discrete problems. The original global PSO is summarized as follows:

1. Initialize an array of particles with random positions and velocities. In the case of image features selection each particle is initialized initially by a feature subset that was randomly set.
2. Evaluate the desired maximization function which is set in this work to be the classifier accuracy for each generated feature subset.

3. Compare evaluation function with particle's previous best value (*pbest*) (previous best obtained accuracy and its corresponding feature subset):

   If current value > *pbest* then *pbest* = current value

4. Compare evaluation function with group's previous best (*gbest*) (best accuracy obtained by all the particles and its corresponding feature subset):

   If current value > *gbest* then update *gbest*,

5. Change particle's velocity (the particle's selected features subset) by:

$$v_{i,j} = v_{i,j} + c_1 r_1 (pbest_{i,j} - p_{i,j}) + c_2 r_2 (gbest_{i,j} - p_{i,j}), \quad \textbf{5-12}$$

   Where $0 \le r_1, r_2 \le 1$ are uniformly distributed random variables (acceleration constants), $0 < c_1, c_2 < 4$ are learning factors. It is recommended in [128] that

6. Move to the new position (new feature subset) $p_{i,j} = p_{i,j} + v_{i,j}$

7. Loop to step 2 and repeat until either a predefined criterion is met or the maximum number of iterations is reached. In this work the criterion is the best classification accuracy (all particles should converge to the maximum accuracy).

5.6.1.2 FS using PSO

The algorithm used in our application is a slightly modified version of the PSO algorithm to fit the binary nature of the feature selection problem.

In this work each particle is treated as a point in a 2^n dimensional space where n denote the total number of features. Let v_{ij} and p_{ij} denote the j^{th} component of the i^{th} particle velocity v_i and position p_i respectively. At every iteration, the particle position is updated by two "best" particles. The first one (denoted by *pbest*) is the best fitness it has achieved so far. Another "best" value that is tracked by the particle swarm optimizer is the best value (denoted by *gbest*), obtained so far by any particle in the population (global PSO).

We also define the fitness of each particle as recognition accuracy corresponding to the features selected by this particle using a pre-specified classifier (in this work, we use SVMs).

The classification error ε is determined by dividing the number of misclassified regions N by the total number of regions R according to equation 5-13 and the fitness (recognition accuracy) is then determined by equation 5-14.

$$\varepsilon = \frac{N}{R} \quad \textbf{5-13}$$

$$fitness = 1 - \varepsilon \quad \textbf{5-14}$$

In our application, the tailored algorithm is:

1. Initializes a group of random particles with $0 \le p_{i,j} \le 1$, *gbest* =0
2. Update *gbest* that maximizes the classification accuracy
3. Update $v_{i,j}$ using equation 5-12 and further update $v_{i,j}$ using :

$$v_{i,j} = 2\sigma(v_{i,j}) - 1, \qquad \textbf{5-15}$$

Where $\sigma(x)$ is the sigmoid function used to limit the velocity to 1, the function is given by:

$$\sigma(x) = \frac{1}{1+\exp(-\alpha x)} \qquad \textbf{5-16}$$

Where $\alpha = 1$

4. The position $p_{i,j}$ is then updated using the initial position and the velocity obtained from equation 5-15. Since we are facing a binary problem the position is either 0 or 1 means either choose this feature or leave it.

$$p_{i,j} = f_i(p_{i,j} + v_{i,j}), \qquad \textbf{5-17}$$

Where the selected feature subset at each stage is given by:

$$f_i = (T(p_{i,0}), T(p_{i,1}), \ldots T(p_{i,2^n-1})) \qquad \textbf{5-18}$$

$$\text{and} \quad T(x) = \begin{cases} 1, & x \ge 0.5, \\ 0, & x < 0.5. \end{cases} \qquad \textbf{5-19}$$

5. The *gbest* value is then updated to take the value of the best particle in the population.
6. The algorithm stops when either the accuracy is settled at the best possible accuracy or when the algorithm reaches a predefined number of iterations. It was found that the algorithm did get to the optimum accuracy even for small number of iterations.

The output of the PSO FS algorithm is a vector composed of ones that correspond to the selected features and zeros that correspond to the rejected features.

In order to visualize the algorithm convergence, the behaviours of four different particles are examined and shown in Figure 5-10 where particle three converged to the best accuracy fast, particle one followed, particle four was next and finally particle two did converge. Therefore, by the end of the algorithm all the particles should have been arrived to the best accuracy (the best feature subset). It should be noticed that at each run, a different feature subset was obtained, yet all of the selected feature subsets obtained the same classification accuracy. This result proves that different feature combinations have the same information about the tissue class, i.e. each feature might add some

information to the decision making until the final amount of information needed for the best classification is reached.

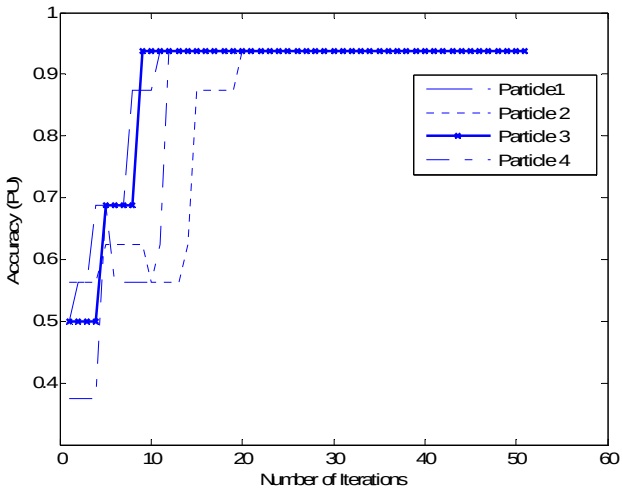


Figure 5-10 Particles' Convergence

Since the data set we are using is not big enough and the PSO doesn't guarantee global optimum, another optimization based FS algorithm is used in this work to support the decision obtained by the PSO. Therefore, Ants Colony Optimization is used in this work to support the PSO result.

5.6.2 Ants Colony Optimization (ACO)

Ants Colony Optimization introduced in [129] is another heuristic optimization method for solving optimization problems. ACO borrows ideas from biological ants. Experiments with real ants showed that almost blind insects like ants go from the nest to the food source and backwards then, after a while, the ants prefer the shortest path from the nest to the food source. A moving ant lays some pheromone thus making a path by a trail of this substance. The ants usually follow trails with higher pheromone. Naturally, larger amount of pheromone will accumulate on the shorter paths to good food sources because larger number of ants will cross it back and forth per unit time as compared to longer paths. Figure 5-11 illustrates the behaviour of natural ants that were moving along a path and then an obstacle was suddenly introduced. The choice of the ants to go either to the right or to the left is determined by the intensity of pheromone trails on each side. A higher amount of pheromone on the right path gives the ant a stimulus to follow this direction.

In this work, a simple algorithm that borrows ideas from the ACO method is presented and it leads to a feature subset with best accuracy.

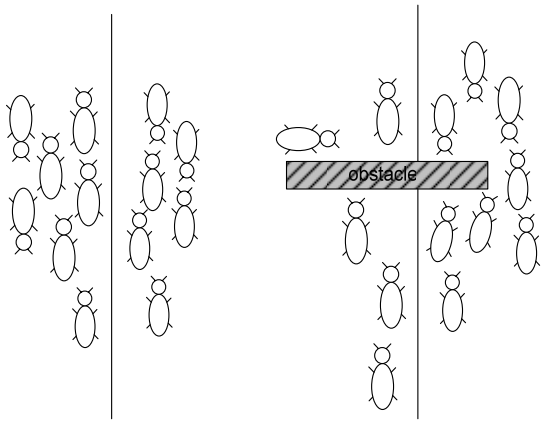


Figure 5-11 illustration of the natural ants moving from the nest to the food

5.6.2.1 ACO for Feature Selection

The ACO is used as a base for the feature selection algorithm used in this section. The system is initialized with a group of ants moving across a full binary tree of depth n and 2^n leaves. Each leaf corresponds to one of the possible 2^n feature subsets. The root of the tree corresponds to the nest of the ants and the accuracy of the classifier based on the feature subset associated with each leaf corresponds to the amount of food found at that food source (classification accuracy in our FS problem).

The algorithm proceeds by iterating through the following basic steps:

- *Construct a solution for all ants*: which feature subset should each ant choose? At the first iteration, all the ants will move randomly (choose random feature subset). However, on subsequent iterations, the ants' choices will be influenced by the intensity of the pheromone trails (classification accuracy) left by preceding ants. The pheromone level reflects the classification accuracy obtained by choosing the corresponding feature subset.
- *Do a global pheromone update* according to the accuracy obtained by following each path.

Figure 5-12 shows a sample of different paths at the end of the FS process. Path 1 is the path with the most amount of pheromone (maximum possible accuracy). This path started with a small amount of pheromone and the pheromone level increases until the program terminates which means that more ants are passing through it due to its better accuracy results (shortest path). Therefore, the feature subset selected by this path is considered as the best feature subset. Path 2 had less amount of pheromone at the end of the program; therefore, it was not selected as the best path. Paths 3 and 4 did have worse classification accuracy meaning that the feature subsets selected by following these paths are not considered informative to the classifier.

As in the PSO method, the accuracy criterion is based on SVMs classifier.

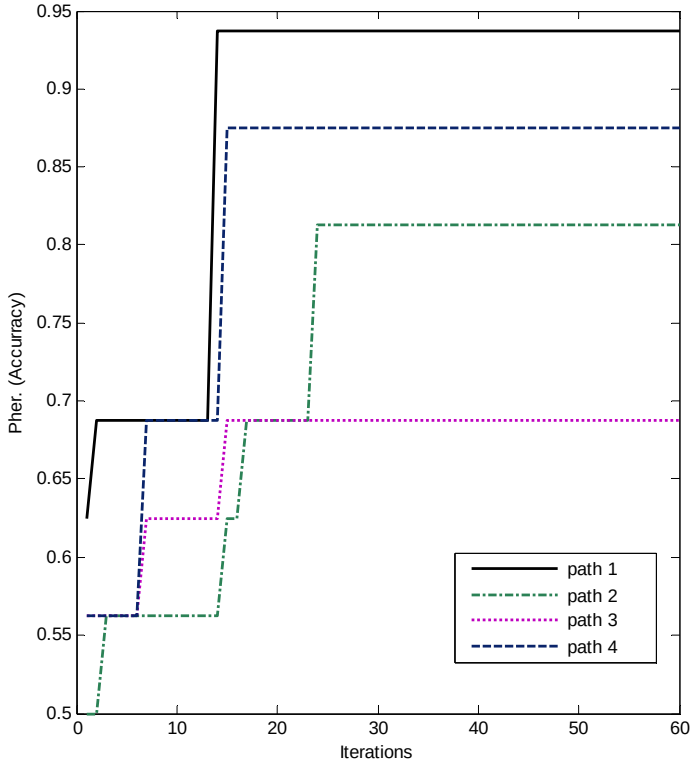


Figure 5-12 Ants Different Paths

5.6.3 Results and Discussion

The classification results are shown in Table 5-4. The results show that the PSO and the ACO obtained the same classification results with 83.33% sensitivity; 100% specificity and 94.4% accuracy (only one misclassified cancer case). The behaviour of some particles is illustrated in Figure 5-10 which shows that the particles usually converge to the best feature subset. Different best solutions can be obtained which proves that each feature might be informative when added to a certain feature sub-set and non- informative when added to another feature sub-set. The ACO paths are also shown in Figure 5-12 where the best path is shown as well as some paths that didn't obtain the best accuracy. Different paths can be chosen as the best solutions, which support the results obtained by the PSO that each feature can add some information to the feature subset until the information needed to obtain the best accuracy is reached. There is no predetermined number of features for either the PSO or the ACO methods.

In the PSO, since particles survive intact from one iteration to the next, then any particle can eventually go anywhere, given enough iterations which means the availability of selecting any feature

subset. In a similar manner, any path can be chosen by the ACO as the initialization process starts with a tree where all possible paths can be explored.

Therefore, either one of the proposed Alife feature selection methods proved their ability to select a feature subset that obtained the same classification accuracy as that obtained using the MIFS. However, MIFS needs a predetermined number of features which is a challenging requirement if we cannot estimate the optimum number of features correctly. Moreover, the MIFS algorithm involves several stages which are the MIFC and MIXY calculations and optimizing equation 5-11 which involves a greedy algorithm. Using a greedy algorithm might cause a nesting problem where a feature that has been discarded cannot be rechosen. Therefore, one of the implemented Alife methods is applied for the next chapter for spectral feature selection as it is less time consuming than the MIFS with the same accuracy results, moreover, Alife methods do not suffer from the greedy problems. Furthermore, the size of the selected feature subset dosen't have to be predetermined. The PSO is elected to select the spectral feature subset in the next chapter.

Table 5-4 Classification Results Using the Alife Methods

PSO		cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			
ACO		cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy			

5.7 Conclusions

In this chapter, different feature selection avenues were tackled. MIFS algorithm was implemented and applied with an almost optimal feature subset that obtained excellent sensitivity, specificity and accuracy. However, using the MIFS involves the choice of the size of the feature subset; MIFS also might suffer from nesting problem due to application of a greedy algorithm for feature selection. Therefore, different Alife based optimization methods are tailored to fit the texture feature selection and were applied to the ROIs constructed features. These methods are the PSO and the ACO. The Alife based FS methods obtained excellent classification results that were the same as the MIFS, but with much less computational time. The SVMs classifier was used for the accuracy evaluation for all the proposed FS methods.

Chapter 6
Spectral Feature Analysis

6.1 Preamble

In the previous chapters of this thesis, several aspects were tackled which are: In chapter 3, the suspicious ROIs were identified from the segmented TRUS images using a novel system that integrates the frequency information, the spatial information as well as the expertise information. In chapter 4, different statistical features were constructed and three different classifiers namely: CNN, DT and SVMs were compared based on the classification accuracies using the constructed statistical features. One of the conclusions that were extracted from chapter 4 recommended the use the SVMs classifier throughout the rest of the thesis. The other conclusion was that the higher order statistics are excluded from further analysis as they supply more information than what can be afforded at this stage provided that the available data is limited by the HVS. While in chapter 5, different feature selection methods were introduced and applied to the constructed statistical TRUS image ROIs' features. The recommendation of chapter 5 was to use any of the introduced Alife FS techniques for the feature sets constructed from the ROIs as the MIFS demanded more computational efforts and it obtained the same results as the proposed Alife algorithms. In this chapter, several novel spectral feature sets are proposed and constructed. These feature sets are implemented and applied for the first time for the application of prostate tissue characterization.

CAD using ultrasound is typically based on constructing parameters from the returned echoes in order to identify the tissue type present in the image. This is usually achieved using either one of two main avenues. The first avenue is extracting parameters from the image texture where features are constructed from the image after the echo processing is performed. This avenue was the focus of chapter 3 in this thesis. The second avenue is constructing parameters from Radio Frequency *(RF)* signal that is recorded from the returned echoes prior to machine processing. The RF avenue is typically applied for prostate tissue characterization using spectral parameters in which the power spectrum of the *RF* signal is linearized and some features such as the y-intercept and the mid-band are extracted from the linearized power spectrum [66, 67, 68].

Using the typical *RF* signal has the advantage of being free from the machine processing distortion. While for the image texture applications, data is easier to collect and implement visually. Different spectral features from the RF-echo signals, i.e., before the ultrasound image formation, have been used successfully for prostate cancer diagnosis. Unfortunately, the RF signal may not be available to most researchers. In this chapter, an innovative technique is proposed to scan the ROIs in order to be able to apply different spectral methods to the TRUS images ROIs in an attempt to get the benefit of the successful RF method when provided only with the TRUS images (RF data is not available). Therefore, the spectral based image feature construction is introduced in this work and the proposed spectral based feature analysis algorithm is shown in Figure 6-1. In the stage that follows the ROI identification explained in chapter 3, different spectral features are constructed from the ROIs. Next,

a classifier-dependent feature selection algorithm is applied to select the most salient and informative features. Two Alife FS methods were explained and applied to the TRUS image texture features in the previous chapter and are expected to have the same selection power for the TRUS spectral features. The PSO, one of the elected Alife based methods that were explained in chapter 5, is used to select a feature subset from the constructed spectral features. Finally, SVMs classifier is used for testing the system's accuracy using the selected feature sets.

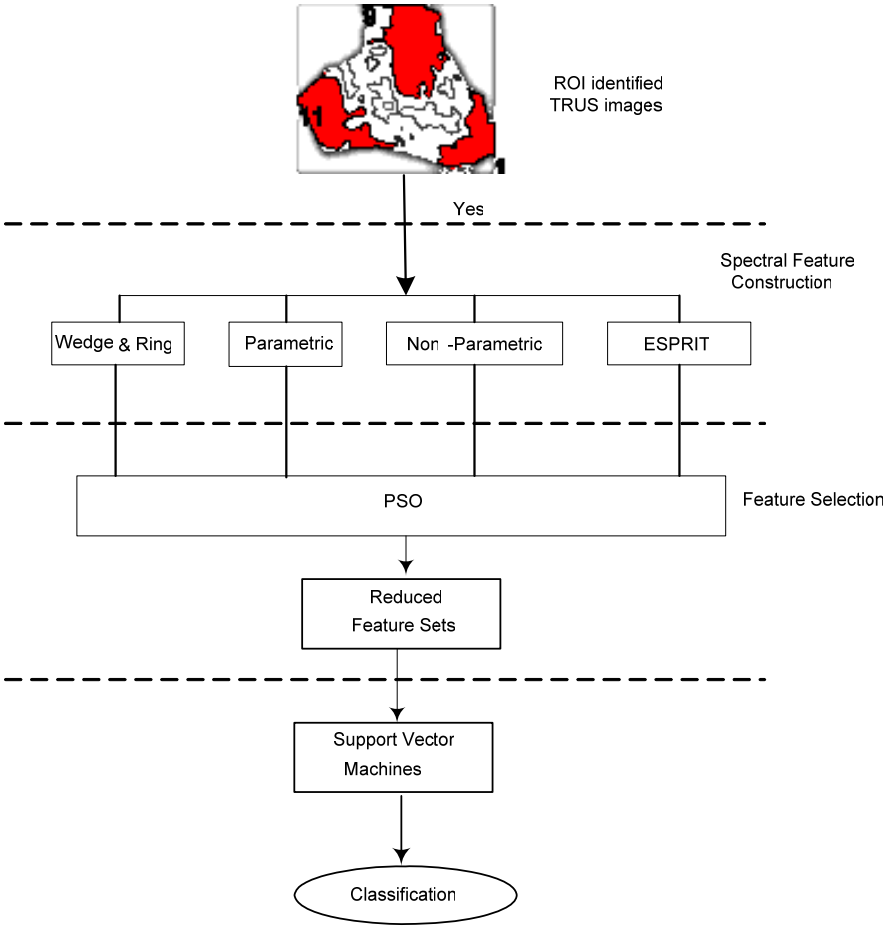


Figure 6-1 The Designed Spectral Based CAD Algorithm

6.2 ROI Signal Generation Results

For the spectral feature construction methods used in this chapter and explained in sections 6.6, 6.7 and 6.8, the pixels within each ROI are aligned to form a 1-Dimensional signal. The pixels alignment method is illustrated in Figure 6-2 and a sample of the obtained signal is shown in Figure 6-3. It is clear that the grey level changes slightly as we move away from the pixel in which the aligning process started until the end of the first row is reached as shown from the figures. The aligning

process then takes pixels from the next row starting from the pixel just below the last pixel in the previous row as illustrated in Figure 6-2. The grey level changes smoothly in the opposite direction until it gets to the last pixel in the second row, where the grey level is very close to the one we started with. This pattern is illustrated in Figure 6-2 and Figure 6-3. The signal follows the same pattern with a slight change between each pixel and its neighboring one until the aligning process is finished by covering the whole ROI.

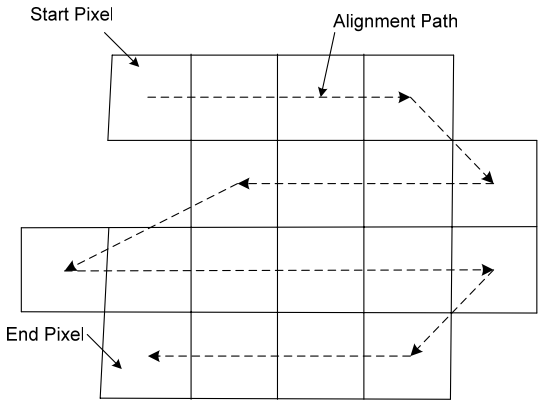


Figure 6-2 The way pixels are aligned to obtain the ROI signals

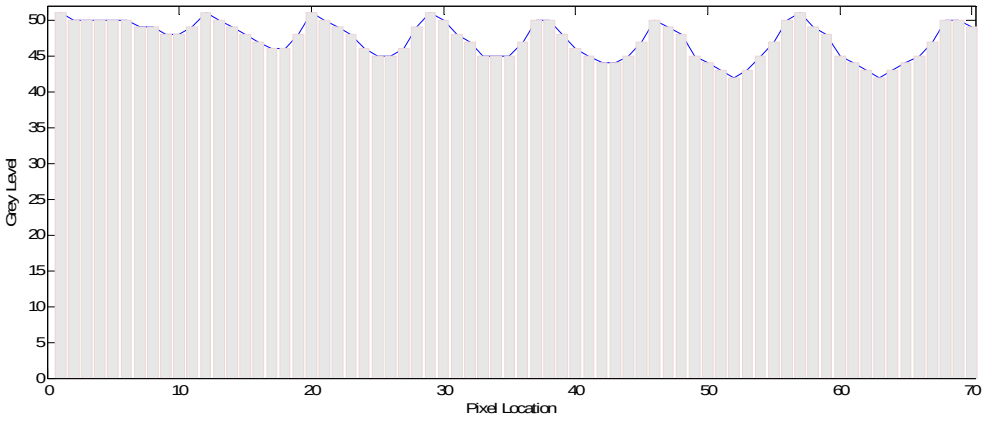


Figure 6-3 Zoom Into the Grey Level Signal of a Cancerous Region

Figure 6-4 shows a sample TRUS image with the doctor's identification highlighted on the TRUS image as well as the proposed algorithm identified ROIs. The signals that correspond to different regions in the image are shown in Figure 6-5, Figure 6-6 and Figure 6-7, where Figure 6-5 and Figure 6-6 show the signals of the two ROIs that correspond to regions identified by both the doctor and the proposed algorithm (regions 3 and 10) which means that the algorithm did capture the doctor's information.. Figure 6-7 shows the signal that corresponds to a region that is identified by neither the proposed algorithm nor the doctor (non-cancer region).

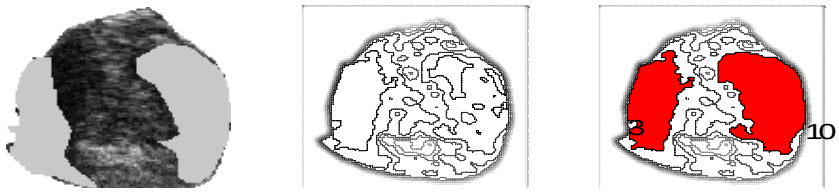


Figure 6-4 Prostate TRUS Image and the Corresponding Identified ROIs

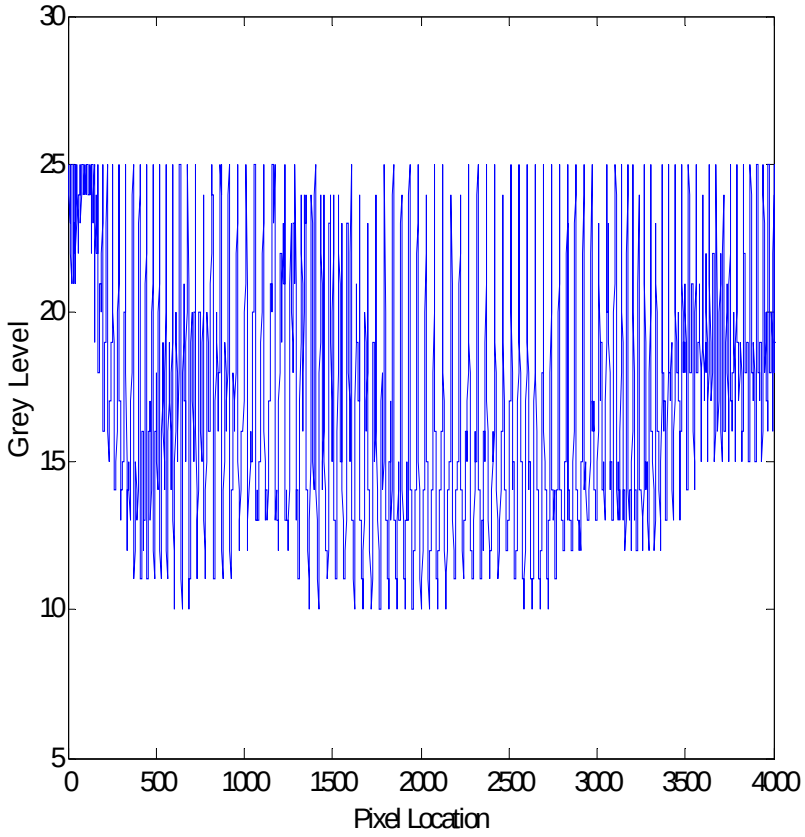


Figure 6-5 ROI Signal for Region 3 (cancer region)

It is clear from the presented figures that the signals corresponding to the cancerous regions look totally different than that signal corresponding to non-cancerous region. The cancerous regions signals' grey level is relatively small compared to the grey level of the non-cancerous regions, which means that the cancerous regions are usually darker. The obtained signals look very rich and they contain lots of information more than just the grey level values. Some of these information can be extracted using signal processing techniques.

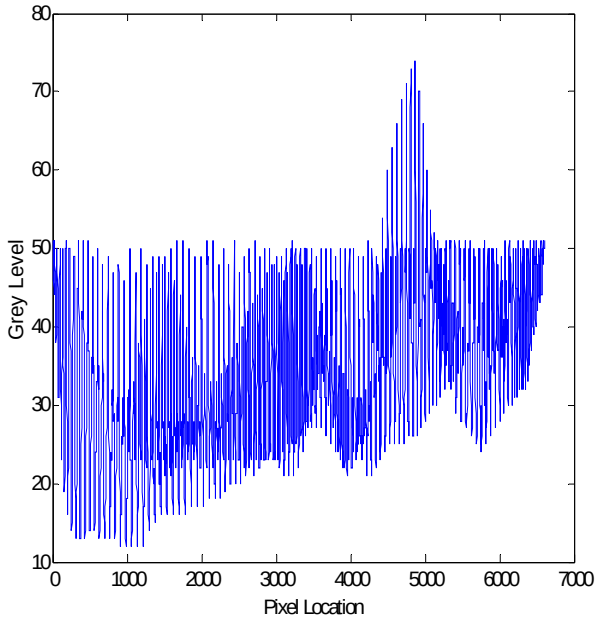


Figure 6-6 ROI Signal for Region 10 (cancer region)

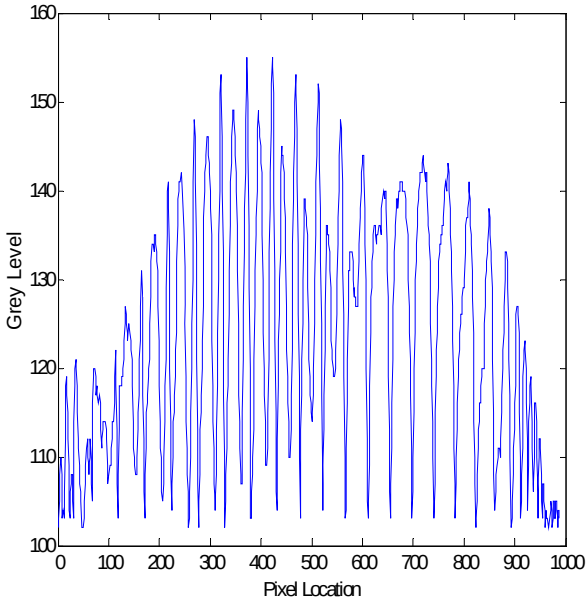


Figure 6-7 ROI Signal for a non-cancer region

Figure 6-8 shows another TRUS sample image with the doctor's identification highlighted on it as well as the proposed algorithm selected ROIs. The signals that correspond to different ROIs are shown in Figure 6-9 and Figure 6-10. Figure 6-9 shows the ROI's signal corresponding to the

doctor's highlighted cancerous region (region 3). Figure 6-10 shows the signal that corresponds to a region that was selected by the filter, as a suspected cancerous region, and missed by the doctor (region 1).

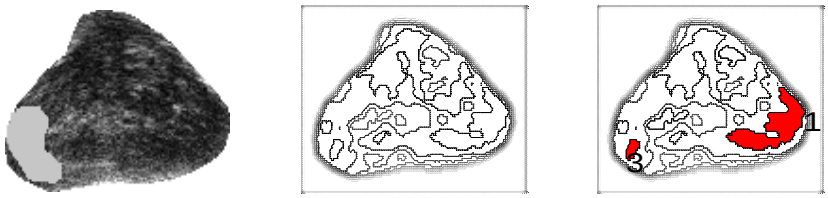


Figure 6-8 Prostate TRUS image and the corresponding identified ROIs

For the TRUS image shown in Figure 6-8, the ROI signal that corresponds to the doctor's marked region (region 3) is shown in Figure 6-9. It resembles the cancer signal shown in Figure 6-5 and Figure 6-6 with a difference in the number of pixels as region 3 in Figure 6-8 is much smaller than the other two regions. However, it does follow the same pattern and grey level values. Moreover, the algorithm did recognize a region whose signal follows the same pattern and grey level values of the cancer regions; however, it was missed by the doctor. Moreover, this analysis proves the need to dig more into the signal and get as much information as possible from it. Furthermore, it allows for the use of the different spectral feature construction methods that were available only for the RF data.

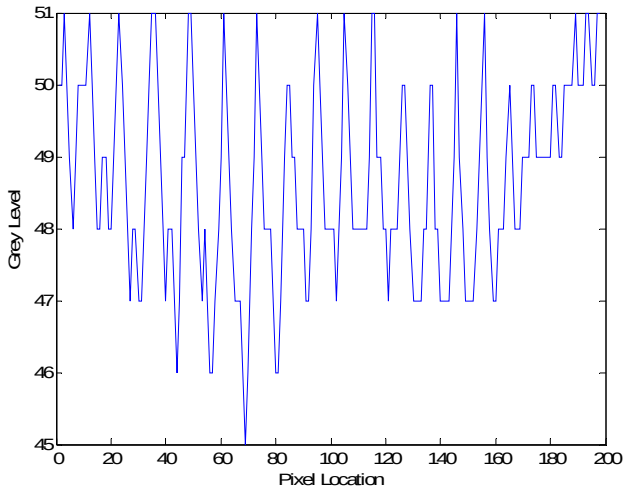


Figure 6-9 ROI Signal for Region 3 (cancer region)

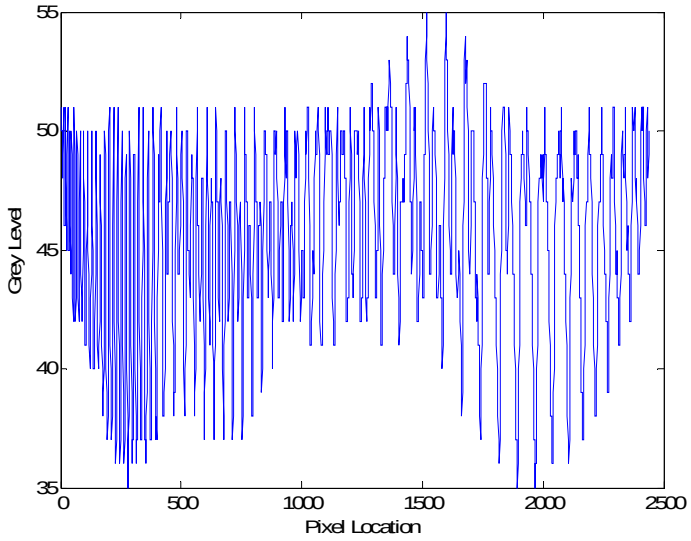


Figure 6-10 ROI Signal for Region 1

Some statistics of the obtained sample signals are calculated and illustrated in the following tables. Table 6-1 shows the mean, median and Standard Deviation StD of the signals that were obtained from the cancerous ROIs that correspond to the doctor's identified regions as well as their corresponding averages. The table shows consistency in the statistics of the cancerous signals.

Table 6-2 shows the same statistics for a ROI that was not highlighted by either the doctor or the proposed ROI identification algorithm (non-cancer region).

Table 6-1 Statistics of Cancerous Regions Signals

	Figure 6-5	**Figure 6-6**	**Figure 6-9**	**Average**
Mean	17	34.12	48	33
Median	16	33	49	32.7
StD	4.16	10.13	4	6.16

The two tables show distinct differences between the statistics of the cancerous and non-cancerous regions signals. This proves the power of the proposed image scanning method that transfers the image from the image format to the signal format. Moreover, the obtained signals are better discriminated visually than the original TRUS image.

Table 6-2 Statistics of Non-Cancerous Regions Signals

	Figure 6-7
Mean	123.5
Median	122
StD	13.6

Table 6-3 shows the statistics of a region whose corresponding signal is shown in Figure 6-10. The region's signal corresponds to a region that was identified by the proposed identification algorithm and was missed by the radiologist. This region has the same pattern and statistics as the identified cancerous regions by both the doctor and the proposed ROI identification algorithm, which proves the out-performance of the proposed ROI identification algorithm that was able to recognize the regions that were missed by the radiologist.

Table 6-3 Statistics of Unidentified Regions Signals

	Figure 6-10 (missed by the radiologist)
Mean	45
Median	46
StD	4

6.3 Spectral Features

Different spectral features are constructed in the following sections from both the TRUS images ROIs and the TRUS images ROIs' signals that were obtained from the images' ROIs as explained and shown in the previous section. The constructed features are categorized to four main categories which are:

- Wedge and ring features (constructed from the ROIs' images) that are explained in section 6.4,
- Parametric Power Spectrum Density (PSD) estimation features explained in section 6.6,
- Non- parametric PSD estimation features explained in section 6.7
- Estimation of Signals via Rotational Invariant Techniques (ESPRIT) features explained in section 6.8.

The last three methods use the ROIs signals for feature construction.

6.4 Wedge and Ring Feature Construction Method

The discrete Fourier transform is the most widely used transform for textured scenes analysis [87]. Information on textural content is provided by constructing features from the Fourier power spectrum, as it represents the strength of each spatial frequency. The textural information within the image can be represented in a more brief form by the transformation of textured images into the Fourier domain. *Coarseness* is one textural property that can be revealed from the power spectrum. Therefore, conclusions about the properties of a texture can be defined [44, 87]. Coarse textures have power spectrum values concentrated near the origin, while fine textures have more high frequency components, hence; have a more spread out spectrum. These observations lead to the features that are constructed from the Fourier domain using ring and wedge filters. To illustrate the PSD representation of a coarse texture, Figure 6-11 shows the PSD for a 32x32 test image that is totally black except for a square white patch of 10x10 pixels located at the center of the image. This image represents a coarse texture where most of the PSD is concentrated at the origin.

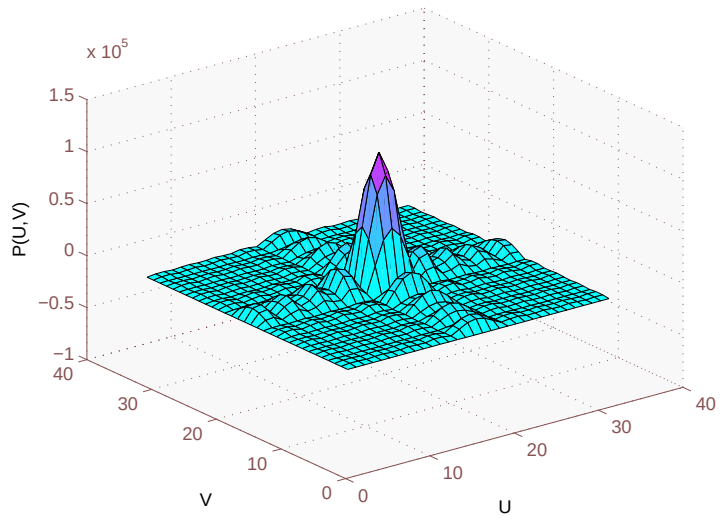


Figure 6-11 PSD of a coarse texture

The wedge and ring filters were introduced in [130] and were used for image segmentation in [131] where these filters were compared to other image features. Typically, the power is measured between rings of inner radius r_1 and outer radius r_2. Moreover, the directionality of the image texture is represented by the average power over wedge-shaped regions centered at the origin of the power spectrum.

This method can be summarized as follows. For any spatial position (x, y) on the image, there is a function $f(x, y)$ representing the grey scale level of the TRUS image. Thus, the 2D Fourier transform pair for an image (assumed to be rectangular) of width M_s and height N_s has the form,

$$\Im\{f(x,y)\} = F(u,v) = \frac{1}{M_s N_s} \sum_{x=0}^{M_s-1} \sum_{y=0}^{N_s-1} f(x,y) \exp\left[- j2\pi \left(\frac{ux}{M_s} + \frac{vy}{N_s} \right) \right] \quad \textbf{6-1}$$

for $u, x = 0,1,2,\cdots, M_s - 1$ and $v, y = 0,1,2,\cdots, N_s - 1$

However, the ROIs are non-rectangular, therefore, zero padding was performed toimbed the ROIs to a rectangular patch.

For deterministic signals, the square of the magnitude of the Fourier spectrum is referred to as the power spectrum of f . This method was applied earlier for liver tissue classification from ultrasound images in [40]. The same method is applied in this work to the TRUS image ROIs and a sample power spectrum that is calculated for an identified ROI is shown in Figure 6-12

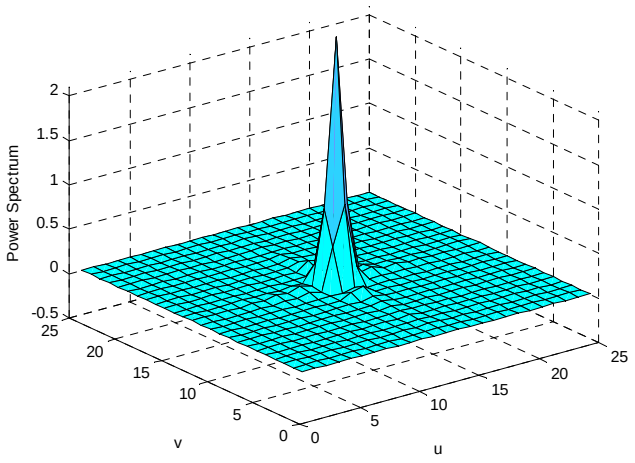


Figure 6-12 Power Spectrum of a TRUS image ROI

Let $P(u,v) = \left|\Im(u,v)\right|^2$ and let $P_{r\phi}(r,\phi)$ denote the corresponding function in the spherical coordinates, i.e., with $r = \sqrt{u^2 + v^2}$ and $\phi = \tan^{-1}(\frac{v}{y})$.

Then the power contained in a ring centered at the origin having inner and outer radii of r_1 and r_2 respectively, is given by $\sum_{r=r1}^{r2} \sum_{\phi=0}^{\pi} P_{r\phi}(r,\phi)$. Similarly, the power contained in a wedge of radius r and bounded between ϕ_1 and ϕ_2 is given by $\sum_{r=0}^{r} \sum_{\phi=\phi_1}^{\phi_2} P_{r\phi}(r,\phi)$.

In this work, the wedge and ring filters method is applied to the ROIs of the TRUS images for the first time, where the energy content in ten wedges and ten rings are calculated and used for TRUS images' ROI tissue classification.

One clear weakness of this method is that rings and wedges geometry do not match the irregular nature of the ROIs (that are not square or rectangular regions) of the TRUS images. To avoid dealing with this irregular image (ROI), the ROIs are converted to a 1-D signal.

For the methods described throughout the rest of this chapter, the pixels within each ROI are aligned to form a 1-Dimensional signal as explained earlier in section 6.2 and different spectral analysis methods are applied to this signal where many spectral features are constructed. This pixel alignment is proposed to take into account all the pixels that are enclosed in the ROI without the approximation by a square that might either add more pixels that don't belong to the ROI or take away some pixels that belong to the ROI.

6.5 Proposed Novel Constructed Features Using the Estimated PSD

In particular, for the case of the estimated PSD, the proposed new feature set that is constructed from the estimated PSD contains features that are different from the ones originally found in literature [67, 68, 66]. In the previous work done in this area, linear interpolation was usually done to the estimated PSD of the RF signals and features were constructed from the straight line. The constructed features in the previous work were mainly the slope, the mid-band PSD and the y-intercept of the straight line. In the proposed algorithm, the estimated PSD was plotted against the frequency for the different ROIs' signals and these plots were visually inspected. A sample of the produced PSD plot is shown in Figure 6-13. The first section that corresponds to smaller frequencies has some information that cannot be easily extracted from this plot. Therefore, the proposed feature construction method suggests plotting the PSD and the frequency on a log-log scale. The proposed plot for one ROI is shown in Figure 6-14 in which the y axis represents -20*log(PSD) and the x- axis represents log(f), while Figure 6-15 represents the estimated PSD for some ROIs. From the above discussion and from the figures, the proposed constructed feature set for each ROI in the proposed algorithm depends mainly on a plot that is similar to that shown in Figure 6-14. In which a similar plot is generated for all the identified ROIs and the features that are expected to carry most of the information in the estimated PSD are constructed from the sample plot that is shown in Figure 6-14. It is clear that the approximation by a straight line that was used earlier in the literature for the RF data is not adequate for our generated ROI signal and will loose lots of information that is present closer to the y-axis. Therefore, the proposed algorithm suggests polynomial interpolation for the first segment and the obtained polynomial coefficients are used as features. A sample of the chosen polynomial is shown in Figure 6-14 which is a very close approximation to the first section in the plot. Moreover, in order not to loose the information in the rightmost section of the plot, the second segment is approximated by a straight line where the slope, the y-intercept are considered as features. Furthermore, some more features are proposed which are:

- The value of the first peak as well as its corresponding frequency,
- The frequencies where the point of maximum curvature occurs as well as its PSD value,

- The slope of the bridge (the 1st ascending section right after the peak) that connects the left and right portions of the curve are also considered as features in the proposed spectral feature set.

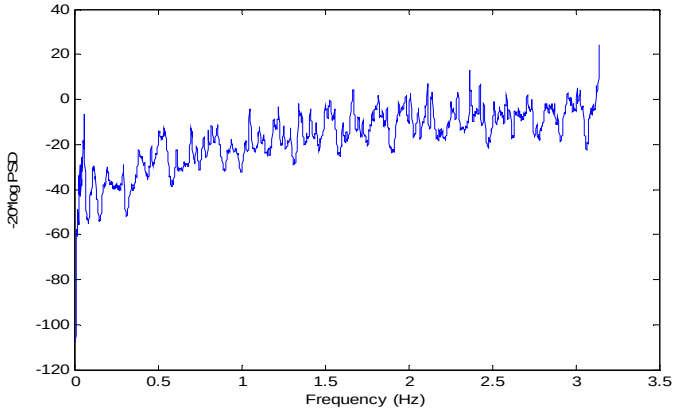


Figure 6-13 The Estimated PSD for a Cancerous ROI

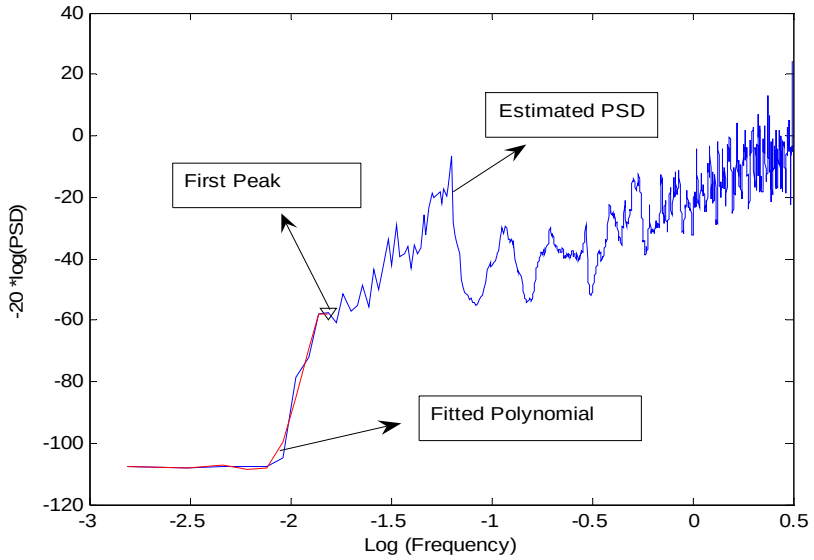


Figure 6-14 The Estimated PSD and the Polynomial Fit to the First Section

Figure 6-15 shows a sample of the estimated PSD for several ROIs which shows that the proposed feature construction method is valid for all the ROIs signals.

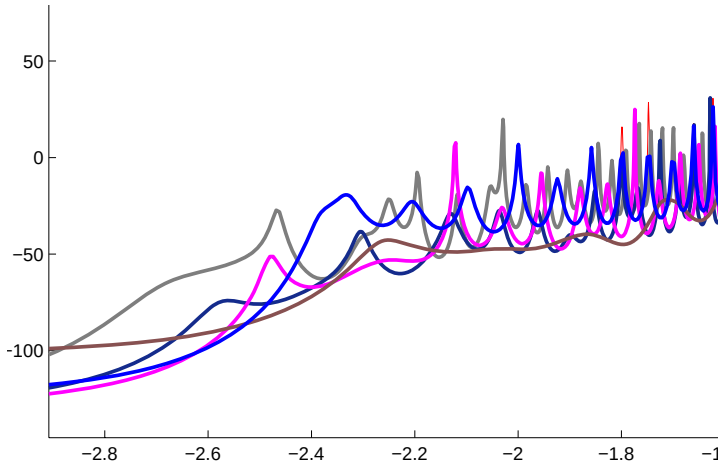


Figure 6-15 Estimated PSD for some ROIs

The used PSD estimation methods have been explained thoroughly in literature [132, 133]. These methods can be categorized into two main categories, namely the parametric and the non-parametric PSD estimation methods.

6.6 Parametric PSD Estimation Methods

In the parametric PSD estimation methods data is typically modeled as the output of a linear system [132, 133] driven by a zero mean white noise, and the parameters of this linear system are estimated. The most commonly used model is the all pole model, whose output is an AR process. Therefore, these methods are referred to as AR methods for spectral estimation. From the mode and the estimated parameters, the PSD can be estimated. The AR methods tend to describe peaky spectral data such as speech signals adequately. AR methods are expected to be suitable for TRUS images, as these kinds of images possess high frequency changes. Therefore, AR methods are selected among the parametric PSD estimation methods that can describe the TRUS images' data in this proposed algorithm. The data is estimated by linear combination of past values (forward prediction) and/or future values (backward prediction). The unknown coefficients are found by minimizing the errors between the predicted values and the original data [132]. All AR models give a PSD estimate given by equation 6-2. [132]:

$$P_{xx}(f) = \frac{\hat{\sigma}^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_p(k) e^{-j 2\pi f k}\right|^2} \qquad \textbf{6-2}$$

Where $\hat{a}_p(k)$ are the estimates of the AR parameters and $\hat{\sigma}^2$ is the estimated minimum mean square value for the p^{th} order predictor. Different parametric PSD estimation methods are used in this work

for the purpose of TRUS images ROI spectral feature construction. These methods are briefed in the following subsections. The classification accuracy using these constructed features is shown in section 6.10.

6.6.1 Yule-Walker AR Method

Also named as the autocorrelation method, as the autocorrelation is estimated from the data and the estimates are used for the autoregression model parameters. This method fits an autoregressive (AR) model to the windowed input data by minimizing the forward prediction error in the least squares sense [132] to estimate the PSD of the input. This formulation leads to the Yule-Walker equations, which are solved by Levinson-Durbin recursion where a biased form of the autocorrelation estimate is recommended which ensures that the autocorrelation matrix is positive definite. Therefore, the matrix is invertible and solution is guaranteed to exist. The AR model parameters are obtained by solving the following Yule-Walker equations:

$$\begin{bmatrix} \gamma_{xx}(0) & \gamma_{xx}(-1) & \vdots & \gamma_{xx}(-p+1) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \vdots & \gamma_{xx}(-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{xx}(p-1) & \gamma_{xx}(p-2) & \vdots & \gamma_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \gamma_{xx}(1) \\ \gamma_{xx}(2) \\ \vdots \\ \gamma_{xx}(p) \end{bmatrix} \quad \textbf{6-3}$$

where $\gamma_{xx}(m)$ represent the autocorrelation sequence that is estimated from the data by:

$$\gamma_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n)^* x(n+m) \qquad m \geq 0 \quad \textbf{6-4}$$

The estimated minimum mean square value for the p^{th} order predictor is given by:

$$\hat{\sigma}^2 = \gamma_{xx}(0) \prod_{k=1}^{p} (1 - |\hat{a}_k(k)|^2) \quad \textbf{6-5}$$

6.6.2 Burg Method

The Burg method estimates the AR parameters using an order recursive least square lattice method, based on minimizing the forward and backward errors in linear predictors, such that the AR parameters satisfy the Levinson-Durbin recursion. The details of the error prediction are given in [132, 134]. This method is reported to obtain better high frequency resolution compared to the Yule-Walker method. It should also yield to a stable AR model. This is the main reason it is applied in this work to capture the high frequencies that exist in the signal accurately. The main disadvantage of this method is that it introduces spurious peaks for high order models.

6.6.3 Modified Covariance Method (MCM)

The covariance method for AR spectral estimation is based on minimizing the forward prediction error. While, the modified covariance method is based on minimizing the forward and backward prediction errors not constrained by satisfying the Levinson-Durbin recursion. It is shown in [135] that the MCM performs better than the Burg AR model for some signals and it is shown in [134] that the modified covariance method performs better than the non-parametric periodogram method for the radar backscatter signals. Therefore it is worth constructing spectral features using the MCM for the TRUS images' ROIs as the ultrasound images are quite close to radar images. The radar images are generated using a similar manner to the ultrasound images. The radar machine sends a microwave beam and records the reflected amount of that beam using an antenna in a similar manner where the ultrasound machine sends an ultrasound wave and receives the reflected wave using its transducer [136].

6.7 Non-Parametric PSD Estimation Methods

Since there is no specific assumption regarding the TRUS image data representation, the non-parametric methods are expected to provide good PSD estimation, as these methods do not need to make any assumptions about how the data was generated. Therefore, non-parametric PSD estimation methods are also applied in this work for spectral feature construction, as they tend to estimate the PSD from the observed signal directly.

Three different non-parametric methods are used in this work. Further details about these methods can be found in [132].

6.7.1 Modified Periodogram

One way of estimating the power spectrum of a signal is to simply find the discrete-time Fourier transform of the samples of the process and take the magnitude squared of the result. This estimate is called the periodogram. The periodogram estimate of the PSD of a length L signal $x_L[n]$is

$$P_{xx}(f) = \frac{|X_L(f)|^2}{f_s L} \qquad \textbf{6-6}$$

Where f_s is the sampling frequency and

$$X_L(f) = \sum_{n=0}^{L-1} x_n[n] e^{\frac{-2\pi j f n}{f_s}} \qquad \textbf{6-7}$$

Since the actual computation of $X_L(f)$can be performed only at a finite number of frequency points N .

$$P_{xx}(f_k) = \frac{|X_L(f_k)|^2}{f_s L}, \quad f_k = \frac{k f_s}{N}, \quad k = 0,1,\cdots,N-1 \qquad \textbf{6-8}$$

In this work the length of the signal that is obtained for each ROI by aligning its pixels is determined and N is chosen bigger than L so that N is the next power of two larger than L. $X_L(f_k)$ is evaluated by zero padding $x_L(n)$ with zeros to the length of N. The original periodogram suffers from the leakage problem that results in large side lobes in the frequency domain, especially for shorter signals which is the case of the ROI signals. The original periodogram is also known to be a biased PSD estimation method where the PSD produced using the periodogram rather represents leaky PSD than the true PSD. Another drawback of the original periodogram is its variance, which does not tend to zero for infinite data points, which means that the periodogram is not a consistent PSD estimator. Therefore, the original periodogram was not used for this work.

The modified periodogram tends to solve the spectral leakage problem that is apparent in the original periodogram. The modified periodogram windows the signal prior to computing the Fourier transforms in order to smooth the edges of the signal. This has the effect of reducing the height of the side-lobes or spectral leakage. This action gives rise to the analysis of side-lobes as spurious frequencies introduced into the signal by the abrupt truncation that occurs when a rectangular window is used. For nonrectangular windows, the end-points of the truncated signal are attenuated smoothly, and hence the spurious frequencies introduced are much less severe. On the other hand, nonrectangular windows also broaden the main lobe, which results in a net reduction of resolution. In the proposed algorithm, a Hamming window is used in order to lessen the effect of large side lobes. Yet, the nonrectangular windows affect the average power of the signal as some of the time samples are attenuated when multiplied by the window.

The modified periodogram estimate of the PSD is:

$$P_{xx}(f_k) = \frac{|X_L(f)|^2}{f_s L U}, \quad \textit{where} \quad U = \frac{1}{L}\sum_{n=1}^{L-1}|w(n)|^2 \qquad \textbf{6-9}$$

and $w(n)$ that is used in this work is a finite length Hamming window.

6.7.2 Welch' Method

Some modifications can be applied to the classic periodogram method that allows for overlapping signal segments as well as windowing the signal segments prior to computing the periodogram. The Welch's method depends mainly on dividing the time series signal into segments that can be overlapping, it then computes a modified periodogram of each segment, and finally the average PSD estimate is obtained [137]. The averaging of modified periodograms tends to decrease the variance of the PSD estimate relative to a single periodogram estimate of the entire data record, which makes it more consistent method for PSD estimation. However, the bias of the Welch's method is reported in [138] to be more than that of the periodogram. Although overlap between segments tends to introduce

redundant information, this effect is diminished by the use of a nonrectangular window, which reduces the importance or weight given to the end samples of segments (the samples that overlap). Therefore, from the above discussion the ROI signal is divided into eight segments (to reduce the variance of the estimated PSD) and a Hamming window (to decrease the redundant information) with an overlap of 50% is used for the ROI PSD estimation using Welch's method.

6.7.3 Multi Taper Method (MTM)

The MTM method uses linear or nonlinear combinations of modified periodograms to estimate the PSD. These periodograms are computed using a sequence of orthogonal tapers [133]. Thompson's multi-taper method (MTM) is supposed to build an improved PSD estimate. Instead of using filters that are rectangular windows such are the ones used in the periodogram method, the MTM method uses a bank of optimal band-pass filters to compute the estimate. These optimal filters are derived from a set of sequences known as Discrete Prolate Spheroidal Sequences (DPSS) [139, 140]. Moreover, the MTM method provides a parameter that could balance the variance and resolution. This parameter is given by the time-bandwidth product, $N \times W$ and it is directly related to the number of tapers used to compute the spectrum. There are always 2N×W-1 tapers used to form the estimate. This means that, as N×W increases, there are more estimates of the power spectrum, and the variance of the estimate decreases. However, the bandwidth of each taper is also proportional to N×W, so as N×W increases, each estimate exhibits more spectral leakage (i.e., wider peaks) and the overall spectral estimate is more biased. In the proposed work, MTM is used to estimate the PSD using a time bandwidth product of four according as suggested in [138].

All the previously explained PSD estimation methods are applied and the features explained in section 6.5 are constructed for each of the previously explained methods. This feature set is then passed to a feature selection technique that specifies the best discriminating spectral features. The results are explained later in this chapter in section 6.9. From the obtained results it was clear that all the proposed feature sets are essential to obtain the best possible accuracy. Hence, there is a need to propose a more compact feature set that attains the important information needed for classification.

Therefore, another novel feature set is constructed for TRUS image analysis for the first time to the best of the author's knowledge. The proposed feature set is the ESPRIT feature set.

6.8 Estimation of Signal Parameters using Rotational Invariant Technique (ESPRIT)

In the proposed algorithm, a novel approach is used for spectral feature construction using TRUS images. Estimation of Signal Parameters via Rotational Invariant Techniques (ESPRIT) appears to be one of the best-known spectral-analysis methods. ESPRIT has a wide application in array signal processing, but the ESPRIT for ultrasound images is rare. ESPRIT has the highest resolution; it has no spectral splatter and is robust in the presence of noise [141]. Unlike Fourier methods that determine the amplitudes, applied to a set of regularly-spaced real frequencies, that best represent the data, ESPRIT answers the more explicit question "What frequencies are present and what are their

amplitudes?." Moreover, ESPRIT can deal with complex frequencies, where the signal damping can be estimated. ESPRIT is used for our proposed system, where the ROIs signals are pixels' with grey level values that are aligned to form a 1-D signal as explained in section 6.2.

ESPRIT was introduced and analyzed in [143, 144]. It is usually applied for high-resolution signal parameter estimation. ESPRIT proved high accuracy and was applied to a wide variety of problems including accurate detection and estimation of cissoids (damped sinusoids). There were several approaches to solve the problem of estimation of signal parameters. These methods started by the maximum likelihood (ML), which was successful but had some limitations like the sensitivity in parameter estimates [142]. A better approach was the Multiple SIgnal Classification (MUSIC) that started by solving the problem neglecting any noise and then applying the effect of noise in a way that obtained a reasonable approximate solution in the presence of noise [144]. ESPRIT is considered more robust (less sensitive) than the MUSIC algorithm. Moreover, it dramatically reduces the computation and storage costs. Therefore, it is much better to rely on ESPRIT in this work to estimate signal parameters from the ROIs lined pixels signal.

The ESPRIT can be explained as follows:

Assume the pixels' signal is in the form of:

$$x(\eta) = \sum_{i=1}^{M} h_i S_i[n] + \eta[n] \qquad \textbf{6-10}$$

$$\textit{where} \quad S_i[n] = \exp(c_i n) \qquad \textbf{6-11}$$

$(c_i = -\sigma_i + j\omega_i)$ is the damping coefficient and frequency of oscillation of the i_{th} signal component, M is the modal order, and h_i is the complex amplitude of the i_{th} signal component.

$$h_i = |h_i| \exp \phi_i \qquad \textbf{6-12}$$

Damping factors and oscillatory transient frequencies are expected to be functions of ROI texture properties and are distinctive characteristics that can be used to differentiate between cancerous and non-cancerous regions.

Consider N samples from the lined pixels' signal, use M consecutive smaller vectors from the N samples.

$$X = [x(n)\, x(n-1) \ldots . x(n-N)]^T \qquad \textbf{6-13}$$

$$\begin{pmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N) \end{pmatrix} = \begin{pmatrix} 1 & \cdots\cdots & 1 \\ e^{-c_1} & \cdots\cdots & e^{-c_M} \\ \vdots & & \\ e^{-Nc_1} & \cdots & e^{-Nc_M} \end{pmatrix} \begin{pmatrix} h_1\, e^{c_1 n} \\ h_2\, e^{c_2 n} \\ \vdots \\ h_M\, e^{c_M n} \end{pmatrix} + \begin{pmatrix} \eta(n) \\ \eta(n-1) \\ \vdots \\ \eta(n-N) \end{pmatrix} \qquad \textbf{6-14}$$

This can be written in a more compact form:

$$X = AS + \eta \qquad \textbf{6-15}$$

Where A can be expressed in two different ways:

$$A = \begin{pmatrix} 1 & \cdots\cdots & 1 \\ e^{-c_1} & \cdots\cdots & e^{-c_M} \\ \vdots & & \\ e^{-Nc_1} & \cdots\cdots & e^{-Nc_M} \end{pmatrix} = \begin{pmatrix} & A_0 & \\ e^{-Nc_1} & . \; . \; . & e^{-Nc_M} \end{pmatrix} \qquad \textbf{6-16}$$

And

$$A = \begin{pmatrix} 1 & . \; . \; . & 1 \\ e^{-c_1} & . \; . \; . & e^{-c_M} \\ . & & \\ . & & \\ e^{-Nc_1} & . \; . \; . & e^{-Nc_M} \end{pmatrix} = \begin{pmatrix} 1 . \; . \; . 1 \\ A_1 \end{pmatrix} \qquad \textbf{6-17}$$

From the last two equations, it is obvious that

$$A_1 = A_0 \begin{pmatrix} e^{-c_1} & 0 & . \; . \; . & 0 \\ 0 & e^{-c_2} & . \; . \; . & 0 \\ . & . & & . \\ . & & . & . \\ 0 & 0 & . \; . \; . & e^{-c_M} \end{pmatrix} = A_0\, \varphi \qquad \textbf{6-18}$$

$$where \quad (\varphi) = diag\left[e^{-c_1} \;\; e^{-c_2} \;\; . \; . \; . \;\; e^{-c_M} \right] \qquad \textbf{6-19}$$

It is required to solve for φ in order to estimate the damping factors and the frequencies (c_i) for different frequency components. However, since A is not available, therefore φ is obtained directly from the data.

Since the ESPRIT algorithm requires the data to be separated into two orthogonal subspaces namely, the signal subspace and the noise subspace. Therefore, the Eigen decomposition of the autocovariance matrix can accomplish this task. Assuming the noise is white stationary:

$$R_{yy} = E\left\{ y(n) \;\; y^*(n) \right\} = A\,\lambda\,A^{*T} + \sigma^2\, I \qquad \textbf{6-20}$$

Where E denotes expectation, R_{yy} the autocovariance matrix of the signals, and σ^2 is the noise variance.

Therefore, define the following matrices of eigenvectors:

Eigenvectors $$E_{sig} = \left[e_1 \quad e_2 \ldots \quad e_M \right] \qquad \textbf{6-21}$$

Noise eigenvectors $$E_{noise} = \left[e_{M+1} \quad e_{M+2} \quad \ldots \quad e_N \right] \qquad \textbf{6-22}$$

And the corresponding matrices of eigenvalues can be given by:

$$(\lambda_{sig}) = diag\ \left[\lambda_1\ \lambda_2 \quad \ldots \quad \lambda_M\right] \qquad \textbf{6-23}$$

$$(\lambda_{Noise}) = diag\left[\lambda_{M+1}\ \lambda_{M+2} \quad \ldots \quad \lambda_N\right] \qquad \textbf{6-24}$$

Therefore, the complete matrices of the eigenvectors are given by:

$$E = \left[E_{sig} \quad E_{noise} \right] \qquad \textbf{6-25}$$

The Eigen decomposition of the autocovariance matrix shown in equation 6-20 can be placed in the form:

$$R_{yy} = E\,\lambda\,E^{*T} = E_{sig}\ \lambda_{sig}\ E_{sig}^{*T} + E_{noise}\ \lambda_{noise}\ E_{noise}^{*T} \qquad \textbf{6-26}$$

A signal subspace E_{sig} in 6-21 can be estimated from the data. Since A and E_{sig} span the same signal subspace, therefore a linear transformation γ that relates A to E_{sig} exists such that:

$$E_{sig}\ \gamma \;\; = \;\; A \qquad \textbf{6-27}$$

If E_{sig} is assumed to be proportional to *A*, this yield:

$$E_0\ \gamma \quad = \quad A_0 \qquad \textbf{6-28}$$

And $E_1\ \gamma \quad = \quad A_1$ **6-29**

And Substitute from 6-28 and 6-29 in 6-18:

$$E_1\ \gamma \quad = \quad E_0\ \gamma\ \varphi \qquad \textbf{6-30}$$

Rearrange

$$E_1 = \quad E_0\ (\gamma\ \varphi\ \gamma^{-1}) \quad = \quad E_0\ \psi \qquad \textbf{6-31}$$

Where $\gamma\ \varphi\ \gamma^{-1} = \psi$ **6-32**

Equations 6-31 and 6-32 are the main equations of the ESPRIT algorithm. Equation 6-31 represents a set of linear equations that can be solved for ψ. Equation 6-32 represents the eigenvalue decomposition of ψ . φ is the diagonal matrix of eigenvalues and γ is the matrix of eigenvectors. Solving 6-31 and 6-32 gives the required solution for φ.

ESPRIT Algorithm

- Define $N+1$ dimensional random vector y pertaining to $N+1$ consecutive data samples $x[0], x[1], \ldots, x[N]$.
- Estimate the autocovariance matrix R_{yy} from the data.
- Compute the generalized eigenvalues and eigenvectors of R_{yy}.
- Estimate the number of components inside the signal *M*.
- Generate a basis spanning the signal subspace and partition it in two different ways E_0 *and* E_1
- Compute the matrix ψ using 6-32
- Compute the eigenvalues of the matrix ψ using (6-33) which represent the poles.

The ESPRIT method was used in the proposed work to estimate the different frequencies that exist in the signal and their corresponding damping factors. The number of modes is chosen to be eight. Therefore, only the frequencies and damping factors that correspond to the first eight modes are considered as the ESPRIT feature set. This way the noise is separated from the original signal. The

output of the ESPRIT feature construction method is eight frequencies for each ROI and their corresponding damping factors. Figure 6-16 shows a sample of the first mode of one of the regions for visualization purposes. The amplitude for this figure is assumed one. From the figure, it is clear that each mode is represented by the ESPRIT as a damped sinusoid whose frequency and damping factor are considered as the features.

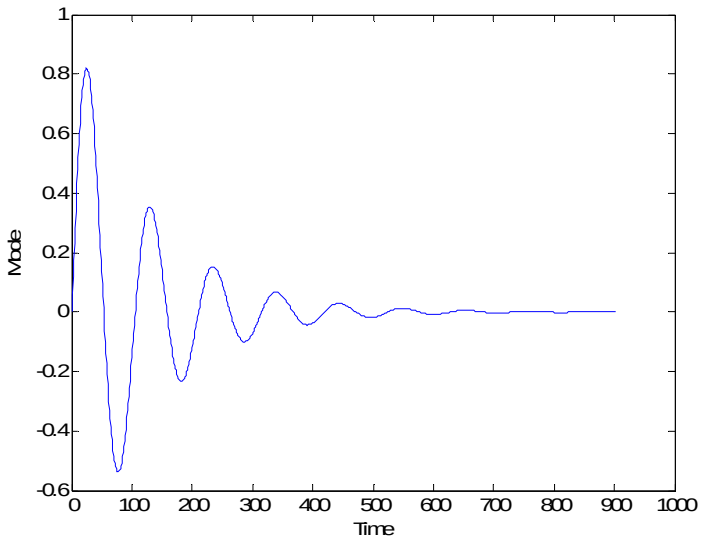


Figure 6-16 A sample first Mode for a ROI signal

The obtained frequencies of the modes that correspond to cancerous ROIs are shown in Figure 6-17, while the frequencies of the modes that correspond to the non-cancerous ROIs are shown in Figure 6-18 for visualization purposes.

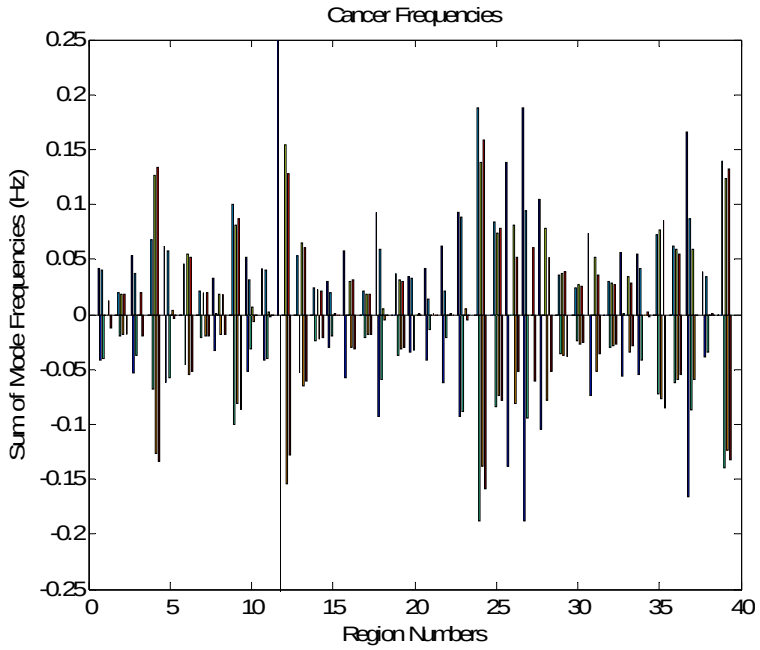


Figure 6-17 Cancer ROIs Mode Frequencies

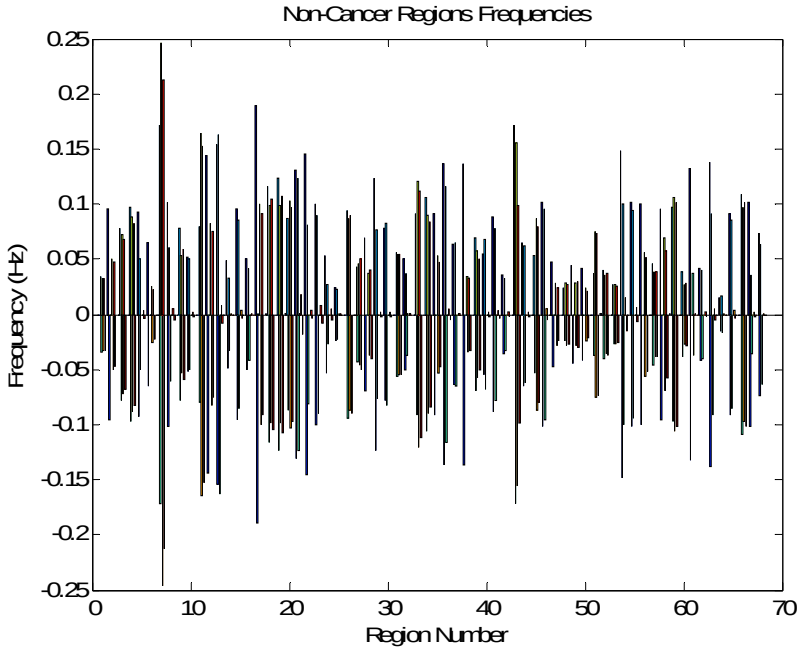


Figure 6-18 Non-Cancer ROIs Mode Frequencies

Figure 6-19 and Figure 6-20 show the histogram of the frequencies that exist in both the cancerous regions and the non-cancerous regions respectively for visualization purposes. It is clear from the figures that the frequency content and pattern are different in the cancerous ROI signals from the non-cancerous ROI signal, which is an indication that the ESPRIT features can be considered as strong features. The frequencies and the corresponding dampings are used as features for the purpose of ROI tissue classification. The size of the ESPRIT feature set is further decreased by applying a feature selection method that chooses the best discriminating features and this is explained in the following section.

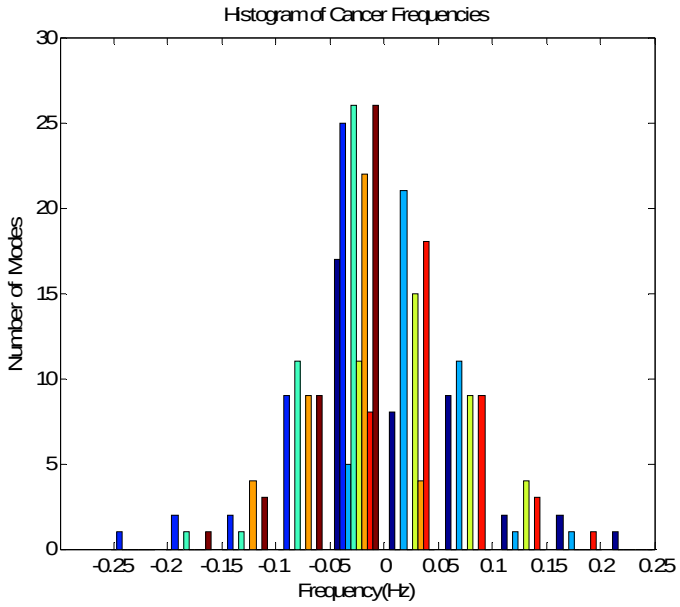


Figure 6-19 Histogram of Cancer Frequencies

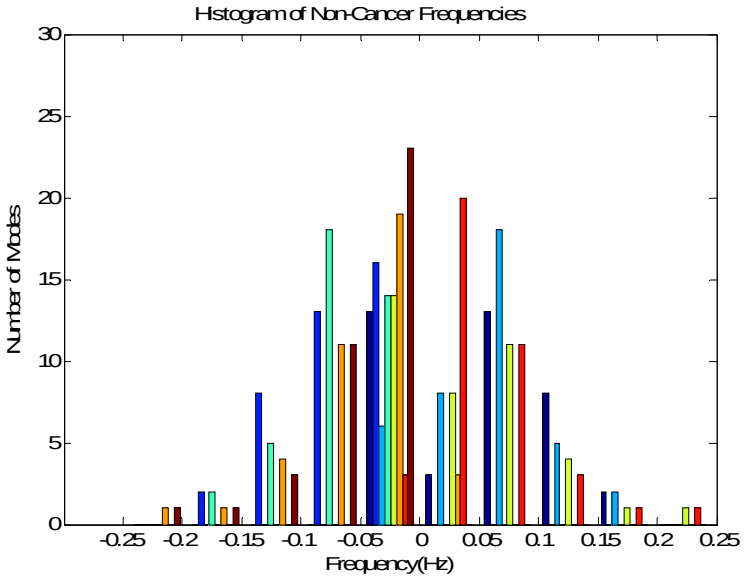


Figure 6-20 Histogram of Non-Cancerous Frequencies

6.9 PSO FEATURE SELECTION

The output of the spectral feature construction algorithms is a group of novel spectral feature sets that need to be analyzed and tuned to identify the features with higher discriminatory power (that carry more information about the ROI tissue type). The wedge and ring method obtained 10 different ring energies as well as 10 different wedge energies. The parametric and non-parametric methods produced feature sets that correspond to each one of the PSD estimation method. These spectral features were explained earlier in sections 6.6 and 6.7. The ESPRIT method obtained eight different frequencies and their corresponding damping factors.

It is expected that some of these constructed features may not be influential and might carry redundant information, which may confuse the classifier. Thus, feature selection and dimensionality reduction is a crucial step to get the best out of the constructed features. The results obtained from chapter 5 recommended one of the Alife FS techniques to be used with the features constructed from the TRUS images.

In this chapter, PSO, the recently proposed optimization technique that was explained and discussed in section 5.6.1 is adopted and used to select an optimal subset from the above constructed features. As mentioned in the previous chapter the function to be optimized is the classifier accuracy.

Finally, the SVMs classifier is used to judge the selected features and obtain the classification accuracy as well as the sensitivity and specificity of the selected feature subsets.

6.10 Classification Results and Discussion

A set of 108 regions has been implemented in this study where 90 regions were used as a training set and the remaining set of 18 regions was used as the test set, as was implemented in the previous chapters. An algorithm was introduced and developed in this chapter for TRUS image scanning in order to form a 1-D signal. This algorithm allows for the adoption of the RF signal processing techniques for TRUS images' signals. This is considered a significant addition that adds on to the existing power of the CAD systems applied on TRUS images. The algorithm was explained in section 6.2 and the obtained signals are shown in Figure 6-5, Figure 6-6, Figure 6-7, Figure 6-9 and Figure 6-10 of this chapter. Moreover, two novel spectral feature sets were introduced and explained in sections 6.5 and 6.8. the classification results obtained utilizing those feature sets are shown in the following subsections.

6.10.1 Spectral Feature Subsets

A subset of each of the different spectral feature sets explained and discussed in this chapter was obtained using the PSO FS and then examined using the SVMs classifier. The accuracy results are shown in Table 6-4. It is clear from the results that the wedge and ring filters possess the least accuracy. As mentioned earlier in section 6.4, this can be explained by the irregular nature of the ROIs (non-rectangular). This also proves that dealing with the regions' pixels as individual data points and treating the ROI as a signal is more appropriate for this application considering the ROIs irregular shapes. The modified periodogram method obtained the least accuracy among the PSD estimation techniques and this can be explained by the effect of spectral leakage that occurs for short data sets. Further, the Hamming windows broaden the main lobe, which results in a net reduction of resolution. Moreover, the variance does not tend to zero as the data length tends to infinity. In other words, the modified periodogram is not a consistent estimator of the PSD especially for short data sets. The Welch's method obtained good results, as it tends to decrease the variance of the estimate relative to the periodogram estimate. Although the overlapping between segments tends to introduce redundant information, this effect is diminished by the use of a Hamming window, which reduces the importance given to the end samples of segments. In the parametric methods, as expected, the modified covariance method obtained the best results. Naturally, combining all the features extracted from the different PSD estimation methods produced the best results as each set has its own information that complements each other.

Table 6-4 Classification results using the spectral feature subsets

Input to the feature selection algorithm	%Sensitivity	%Specificity	%Accuracy	Remarks
Wedge & Ring	66.7%	75%	72.2%	Least accuracy (irregular ROI shape)
Modified Periodogram	66.67%	83.33%	77.7%	Not good (leakage PS, lower resolution)
Multi Taper	66.67%	91.6%	83.3%	
Welch's	83.3%	91.6%	88.89%	Better results (better variance, reduced redundant information)
Yule-Walker features	83.3%	83.3%	83.3%	
Burg features	83.3%	83.3%	83.3%	
Modified Covariance	83.3%	91.6%	88.89%	The best among parametric as expected
All PSD features	83.3%	100%	94.4%	Combined sets obtained the best accuracy

6.10.2 ESPRIT Feature Subset

The ESPRIT feature subset obtained from the PSO FS is also examined using SVMs and the results are shown in Table 6-5. It is clear that the ESPRIT feature subset obtained very good Sensitivity and excellent Specificity, which proves the power of the ESPRIT feature subset in cancer recognition. This was expected beforehand when plotting the frequency histograms that were shown and explained in section 6.8.

Moreover, the classification results that are obtained using only the ESPRIT feature set are as good as those obtained using all the estimated PSD estimation methods. This leads to a conclusion that the ESPRIT features are more adequate and robust to this application, as it did not require the construction of any more features, which reduces the computational cost and increases the accuracy of the proposed recognition algorithm, as there is no need to construct more features while only one feature set obtains the same results.

Table 6-5 Classification results using TLS-ESPRIT feature subset

TLS-ESPRIT feature subset	cancer	Non-cancer
cancer	5	1
Non-cancer	0	12
83.33% Sensitivity; 100% Specificity; 94.4% Accuracy		

6.11 Conclusions

Using spectral features from TRUS images, a novel system for Prostate tissue characterization was introduced. First, the wedge and ring method was used to construct features that represent the energy content in each ring and wedge for each identified ROI. The results using this method were not very good, as the ROIs do not have a regular shape.

Therefore, to avoid the problem of the irregular ROI shape, another approach was proposed and implemented that treats the ROIs as signals. One of the main advantages of the proposed system is the integration of the ROIs pixels to form a signal that allows different signal processing methods to be applied. Second, The ROI signal formation generalized the ROI image format to signal format that allows it to be treated like the RF signal. The proposed ROI conversion to signal method proved to be excellent regarding the application and the obtained signals from the cancer ROIs look very much different from those obtained from the non-cancer ROIs. Different parametric and non-parametric spectral methods for PSD estimation methods were used for the ROI signal analysis. A novel spectral feature set was constructed for each ROI signal after a thorough examination of the estimated PSD plot. The constructed ROIs signal feature sets were examined using the PSO FS and the best discriminating features were selected. When combining all the estimated PSD features, and using the PSO FS as a feature selection algorithm, SVMs classifier achieved a classification accuracy of 94.4%.

Moreover, a novel feature set was constructed for each identified ROI, which is composed of the ESPRIT features that is considered an accurate spectral feature set. The PSO was also used as a feature selection method that is applied to the ESPRIT feature set. The selected feature subset is then tested using the SVMs classifier. The proposed algorithm obtained excellent results with 83.33% sensitivity, 100% specificity and 94.4% overall accuracy.

In conclusion, using the ESPRIT features obtained the same results as using all the other PSD estimation based spectral features, therefore, constructing only one feature set, which is the ESPRIT, is recommended as it saves time and effort while obtaining the same accuracy results for the set of images used in this work.

Chapter 7

LLE Based FD and LEMs Features Dimensionality Reduction

7.1 Introduction

This Chapter focuses on the construction and analysis of a different set of prostate ROIs features for tissue characterization. In Chapter 4, different second order statistical features were constructed, compared and statistically analyzed in order to obtain better classification for prostate ROIs. Moreover, higher order statistical features were constructed and applied for the first time in the context of prostate tissue classification. While in Chapter 5 different novel feature selection methods were introduced, implemented and applied for prostate texture features in order to select the most informative features that possess the highest discriminatory power. Furthermore, in Chapter 6 different novel spectral feature construction methods were introduced and applied to the TRUS images' ROIs. The proposed spectral features are based mainly on converting the TRUS ROI signal to 1-D signal and construct different features that were not applied earlier for the case of TRUS images data. In this Chapter, as briefed in Figure 1-1, from each ROI, two novel different feature sets are constructed; the first set includes the Fractal Dimension (FD) obtained using three different methods and the second set consists of the Laws Energy Measure (LEMs) features. To the best of the author's knowledge, these new textural features are constructed for the first time to serve the prostate tissue characterization problem. To remove the redundancy among the constructed features as well as to overcome the curse of dimensionality, Feature Extraction (FE) is utilized. Feature extraction is recommended in this chapter, as the FD is calculated using three different methods with different approximations which makes these features almost the same, therefore, transferring the data to another feature space allows better data interpretation without any information degradation. Therefore, the recently developed Locally Linear Embedding (LLE) technique is applied to the constructed set of features in order to produce a lower dimensional, neighborhood preserving embeddings of the high dimensional data. Figure 7-1 shows the proposed feature analysis method used in this chapter.

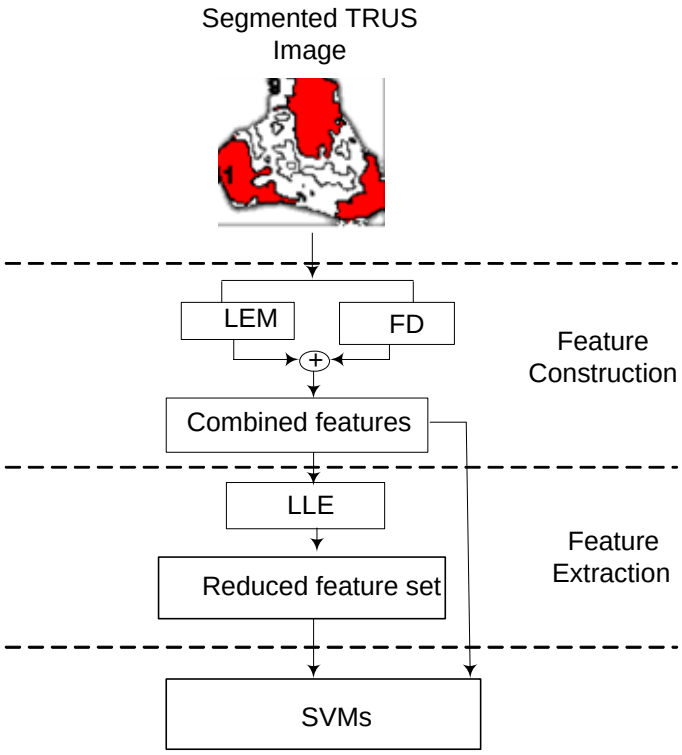


Figure 7-1 Proposed System for Prostate Cancer Diagnosis from TRUS images

7.2 FD Estimation

The most familiar geometrical system is the Euclidean geometry, which deals with regular shapes that have parameters expressed in simple algebraic formula. This system has three dimensions; an idealized line with length but no width has a dimension of one, a planar area has a dimension of two, and a solid volume-occupying object has a dimension of three. A pure fractal object has a complex boundary and the level of complexity remains the same at increasing magnification. The deficiencies of Euclidean geometry in describing the boundary of such objects become apparent when one considers measurements such as perimeter. If the perimeter of the fractal object is to be measured at a certain magnification, then the complexities seen at a closer view would not be included, because they are not visible at the former magnification level and so the estimate of the perimeter would be reduced. This is clear from Figure 7-2 and Figure 7-3 where the first figure shows the image from a further zoom than the second figure. Therefore, the Fractal Dimension (FD) is usually used to describe those objects.

Figure 7-2 The Julia Image under Magnification Level 1

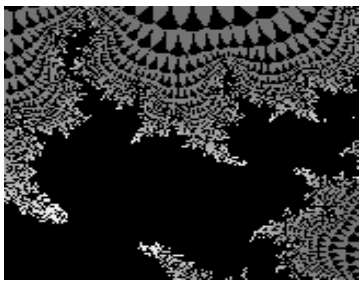


Figure 7-3 The Julia Image under Magnification Level 2

The main problem is that for some structures the level of complexity remains the same at all magnifications, so an absolute measurement of parameters such as the perimeter or area is not possible. The solution to this was to plot the measured perimeter at different magnifications on a log–log graph, which usually results in a straight line. Subsequently, the gradient of that line is known as the Fractal Dimension (FD). FD is different from the Euclidean dimensions in that it does not have to be an integer; its value lies between the topological (Euclidean) dimension of the object (in this case, a one-dimensional line) and the object in which it is embedded (in this example, a two-dimensional plane). The fractal dimension is an index of the space-filling properties of an object so that the closer the dimension is to the topological dimension in which the object is embedded, the greater its space-filling properties. Thus, for example, the coastline of Britain, with a fractal dimension of 1.25, is more space filling than a straight line, because its dimension is greater than one; but it does not completely fill the plane in which it is embedded, because its dimension is less than two.
The fractal dimension is greater for rougher or more irregular structure. The fractal dimension also leads to the definition of a fractal object, which is an object whose fractal dimension is greater than its topological dimension.

Fractal analysis has proved to be useful in describing the irregular surfaces by a single measure [146]. The fractal dimension, describes the degree of irregularity in the surface texture. The rougher or more irregular the structure, the greater its fractal dimension is. Fractals are commonly used to represent natural surfaces such as coastlines and images. Histopathologists recognized the similarity between these images and the boundary of a malignant tumour, such as a breast carcinoma, when viewed by light microscopy. From low to high magnification, the level of complexity of the boundary of the tumour remains relatively constant down to the level of individual cells [145].

Fractals have been used successfully for ultrasound image analysis, where it was among the features used for fetal lung maturity [101]. Assuming that the ROI is a 2-D surface allows the fractal analysis to capture the irregularities in the ROI boundary only. Therefore, to illustrate the TRUS image ROI irregularity, a sample of a sample of a ROI, where the pixel value is represented by the height above the plane, is shown in Figure 7-4. From the figure it is clear that the ROI can be better considered as an irregular 3-D surface not only an irregular 2-D boundary. This method in interpreting the ROIs is introduced in this work for the first time for the TRUS images. It extends the analysis of the ROI to a 3-D surface which allows for capturing more information as it takes into consideration the irregularity in the surface not only the boundary. Measuring the irregularity of the ROI surface is expected to form informative feature as the irregularity of the cancerous regions is expected to be different than the irregularity of non-cancerous regions. Therefore, the FD is investigated in this work for prostate tissue classification.

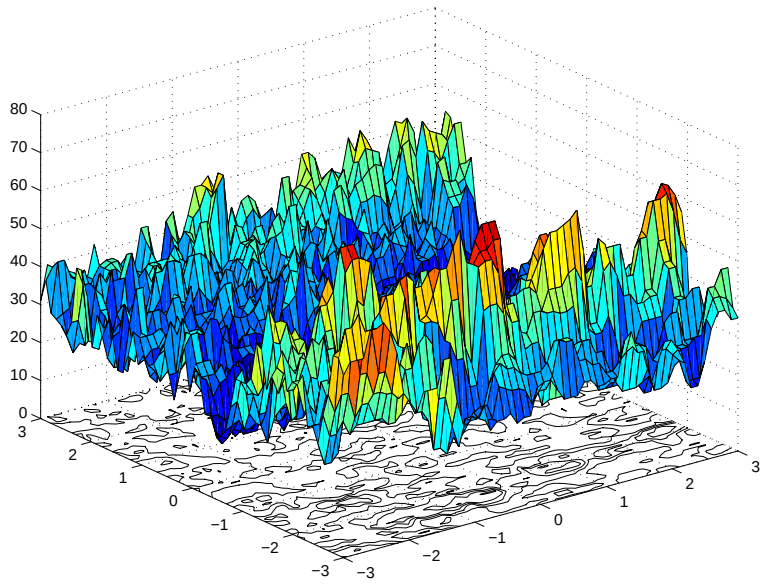


Figure 7-4 TRUS image with pixel values represented as the height above the (x, y) plane

Three different FD estimation methods are used in this chapter. The first FD estimation method is the box counting technique; the second is the Differential Box Counting (DBC) while the third is based on the Fractal Brownian motion (FBm) which uses the Power Spectrum Density (PSD) for FD estimation.

7.2.1 The Box-Counting Fractal Dimension:

The box counting proposes a logical way of characterizing the roughness of structures when no self-similarity is apparent. The box-counting dimension FD_b of any surface may be formally explained as follows [146, 147, 148]:

Subdividing the surface into a lattice of grid size $r \times r$, where r is being reduced, then it follows that N_r is the number of cubes that are contained in the surface for each value of r. The equivalent definition of FD_b is:

$$FD_b = \lim_{r \to 0} \frac{\log N_r}{\log (1/r)} \quad \textbf{7-1}$$

This implies that the box-counting dimension FD_b and N_r are connected by the following power law relation:

$$N_r = \frac{1}{r^{FD_b}} \quad \textbf{7-2}$$

This can be proved by taking logarithms of both sides of equation 7-2,

$$\log N_r = \log\left(\frac{1}{r^{FD_b}}\right) \quad \textbf{7-3}$$

Rearranging using logarithmic properties,

$$\log N_r = FD_b \log\left(\frac{1}{r}\right) \quad \textbf{7-4}$$

From which it follows that,

$$FD_b = \frac{\log N_r}{\log(1/r)} \quad \textbf{7-5}$$

From equation 7-5 it is possible to make an analogy to the equation of a straight line, $y = mx + c$, where m is the slope of the line and c is the y intersection.

Therefore, using this method, the FD is typically estimated by placing each ROI, onto a grid having boxes of size r × r filling up the complete grid [147]. The grid boxes containing some of the structure are then counted. The total number of boxes in the grid, which contain some of the structure, is N_r. The algorithm continues by altering r to progressively smaller sizes and counting N_r. The next step is to plot log (1/r) against log (N_r) and to fit a straight line through the data. The slope of the line is the box-counting dimension for the structure. In the proposed work, the same method is adopted and the pixel intensity is regarded as the height above the plane, then the intensity level of the image is considered as a rugged surface. The rugged surface is divided to cubes that contain the roof and ground of that rugged surface. Then, the size of the cubes is changed such that for each ROI, r is

varied between one and the minimum difference between the roof of the rugged surface and its floor. It is usually checked that all the box elements are included in the ROI and those boxes are used to calculate the FD. Moreover, instead of counting the number of cubes each time, a feature, which is the variance for each cube, is added up. A sample of the plot of N_r against (1/r) for a cancerous ROI (ROI 3 in Figure 7-5) is illustrated in Figure 7-6 and the corresponding log plot with the linear fit is shown in Figure 7-7. The FD obtained using this method for the sample region is 2.7.

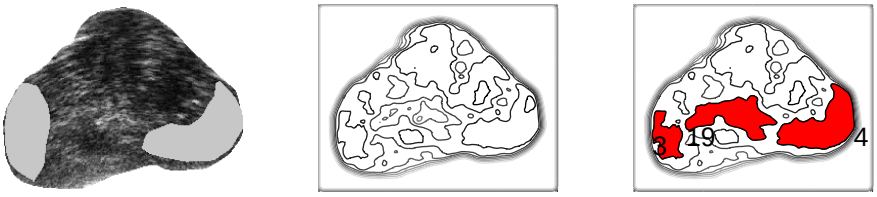


Figure 7-5 TRUS Image used for the illustration of the FD

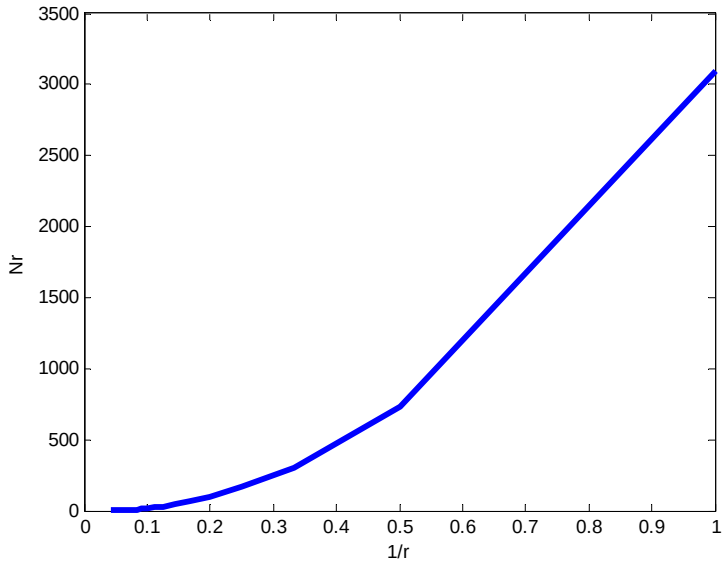


Figure 7-6 The Number of Boxes versus the Box Size inverse for FD Estimation

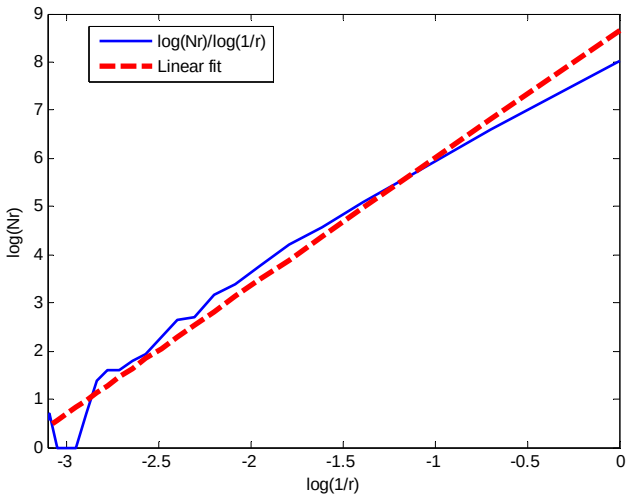


Figure 7-7 The log Plot for the Number of Boxes versus the Box Size for the same ROI

7.2.2 Differential Box Counting method (DBC):

The work done in this section is inspired by the method used in [148]. The basic equation that represents FD was defined in the previous section and is given by:

$$FD = \frac{\log(N_r)}{\log(1/r)}. \qquad \textbf{7-6}$$

Assume that each ROI of $M \times N$ pixels has been divided into $r \times r$ pixels. Next, consider the ROI as a three-dimensional space with (x, y) indicating the position and the third coordinate (z) representing gray-level. The (x, y) space is partitioned into grids of size $r \times r$. On each grid, there is a column of boxes of size $r \times r \times h$. Then, let the minimum and the maximum gray-level of the ROI section located at the $(i, j)^{th}$ grid both fall in the box number k and l as shown in [148].

Then:

$$n_r(i, j) = l - k + 1 \qquad \textbf{7-7}$$

is the contribution of N_r in the $(i, j)^{th}$ grid. Taking contribution from all grids in the ROI, therefore:

$$N_r = \sum_{i,j} n_r(i, j). \qquad \textbf{7-8}$$

Where N_r is counted for differing values of r. Then using equation 7-6, the fractal dimension can be estimated from the linear fitting of $\log(N_r)$ versus $\log(1/r)$.

A sample of the plot for ROI 3 in Figure 7-5 is shown in Figure 7-8 and the log plot is shown in Figure 7-9. The FD obtained using this method for the sample region is 2.81.

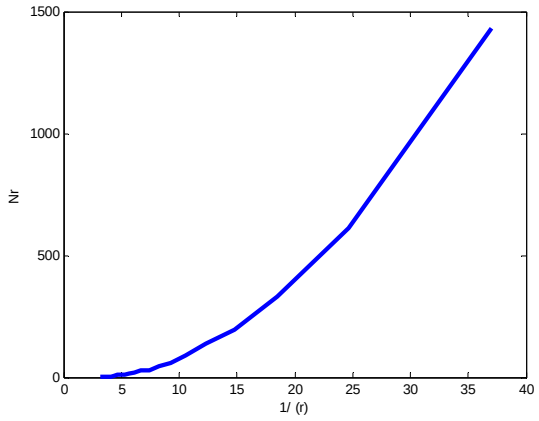


Figure 7-8 Nr versus (1/r)

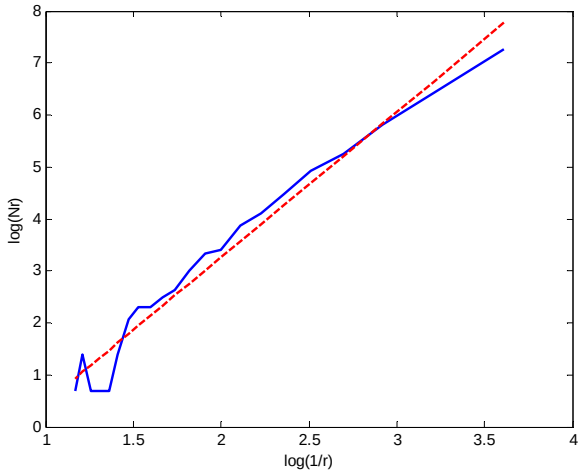


Figure 7-9 log (Nr) versus log (1/r)

7.2.3 Power Spectrum Density (PSD) FD:

This method is based on the Fractional Brownian motion (FBm) model and was explained in [149, 150] and applied in [148] for liver tissue texture from ultrasound images. The FBm was originally used to describe the roughness of natural surfaces. It treats naturally occurring surfaces as the end

result of random walks. An intensity surface of an ultrasound image can be viewed as the end result of a random walk, so the FBm model is expected to suit the analysis of TRUS images' ROIs.

The term random walk is typically used to describe the characteristic action of Brownian motion. This is because the journey of a Brownian particle consists of steps in a seemingly random direction. The displacement of the particle at one time interval is independent of the displacement of the particle at another. Brownian motion is said to be similar under different magnifications and, as such, possesses fractal like properties.

In a strategy to characterize random walks, which could be generalized for fractal processes the **Hurst exponent** was introduced. A Hurst exponent of zero corresponds to a very rough random fractal curve as opposed to an exponent of one, which corresponds to a rather smooth random fractal.

In this work, the power spectrum is used to estimate the FD from the Hurst exponent as recommended in [149]. Where in the frequency domain an average PSD of FBm can be defined and it was shown that the PSD is proportional to $1/f^{\beta}$ with $\beta = 2H + 1$. While in the spatial domain the self similarity structure which underlies the definition of FBm allows to associate to it a FD shown in equation 7-9.

$$FD = 2 - H \qquad \textbf{7-9}$$

Since the FBm has a $1/f^{\beta}$ spectral behavior, then its PSD is supposed to be a straight line in a log-log plot. The slope β of the straight line that fits the PSD is determined and the FD is then estimated.

In this specific application of TRUS image analysis, the PSD of the ROIs' signal that was explained in the previous chapter is estimated and a linear fit is obtained in a log-to-log scale and its slope β is then estimated accordingly. This process is shown in Figure 7-10. Finally, the FD of this region is estimated using equation 7-10 which leads to a FD for the sample region of 2.89.

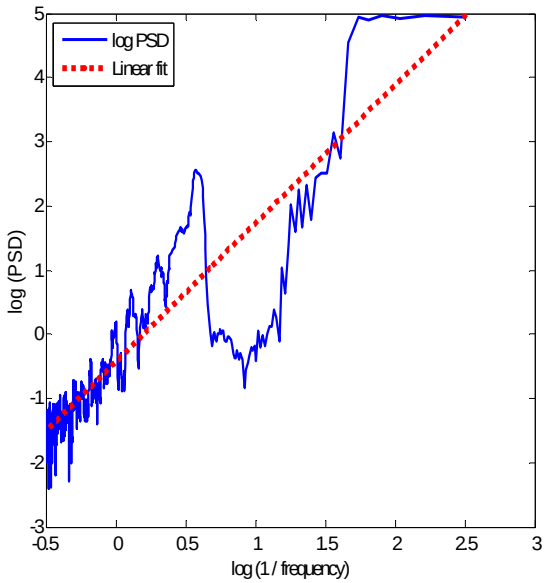


Figure 7-10 FBm FD Estimation

The FD estimation methods explained in this section obtained three different FDs for each identified ROI. In the next section, a different feature set (the Laws Energy Measures) that is constructed from the identified ROIs is explained.

7.3 Laws Texture Energy Measure (LEM)

The next proposed feature set is the Laws Energy Measures (LEMs). The texture energy measures developed by Kenneth Ivan Laws [151] have been used for many diverse applications. These measures are computed by first applying small convolution kernels to a digital image. In the proposed algorithm, LEMs are obtained by convolving Laws masks to the identified ROIs where there are 25 different masks that result in 25 different features for each ROI. Laws' masks are derived from three simple vectors of length 3, which represent the one dimensional operations of center-weighted local averaging, symmetric first differencing for edge detection and second differencing for spot detection.

$$\begin{aligned} L_3 &= (1,2,1), \\ E_3 &= (-1,0,1), \\ S_3 &= (-1,2,-1), \end{aligned} \qquad \textbf{7-10}$$

When these vectors are convolved with themselves or with each other, we obtain five vectors of length 5 as follows:

$$
\begin{aligned}
L_5 &= (1,4,6,4,1),\\
S_5 &= (-1,0,2,0,-1),\\
R_5 &= (1,-4,6,-4,1),\\
E_5 &= (-1,-2,0,2,1),\\
W_5 &= (-1,2,0,-2,1).
\end{aligned}
\qquad \textbf{7-11}
$$

Where L_5 performs local averaging, E_5 and S_5 are spot and edge detectors, R_5 and W_5 are "ripple" and "wave" detectors. Laws' masks are obtained when each of the row vectors of length 5 is multiplied by each column vector of the same length. The obtained masks are shown in Appendix A and two sample masks are:

$$
L_5{}^t E_5 = \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ -4 & -8 & 0 & 8 & 4 \\ -6 & -12 & 0 & 12 & 6 \\ -4 & -8 & 0 & 8 & 4 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix} \qquad \textbf{7-12}
$$

$$
L_5{}^t S_5 = \begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ -4 & 0 & 8 & 0 & -4 \\ -6 & 0 & 12 & 0 & -6 \\ -4 & 0 & 8 & 0 & -4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix} \qquad \textbf{7-13}
$$

These masks are generated and convolved with each of the ROIs and the resulting images' variances are used as ROIs' texture features. A sample of the Laws masks in the frequency domain in shown in Figure 7-11. It looks like this specific mask acts as some kind of a band pass filter. However, there are 25 different masks that filter the ROI in different ways to obtain feature images.

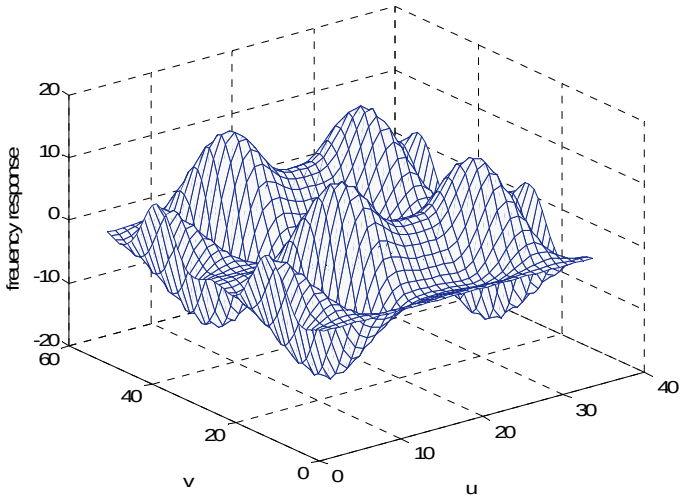


Figure 7-11 A Sample of the Laws Masks' Frequency Response

One sample ROI (region 3 in Figure 7-5) is shown with some feature images in Figure 7-12, Figure 7-13, and Figure 7-14 where the original ROI was convolved with different Laws masks and the feature images are shown in the figures.

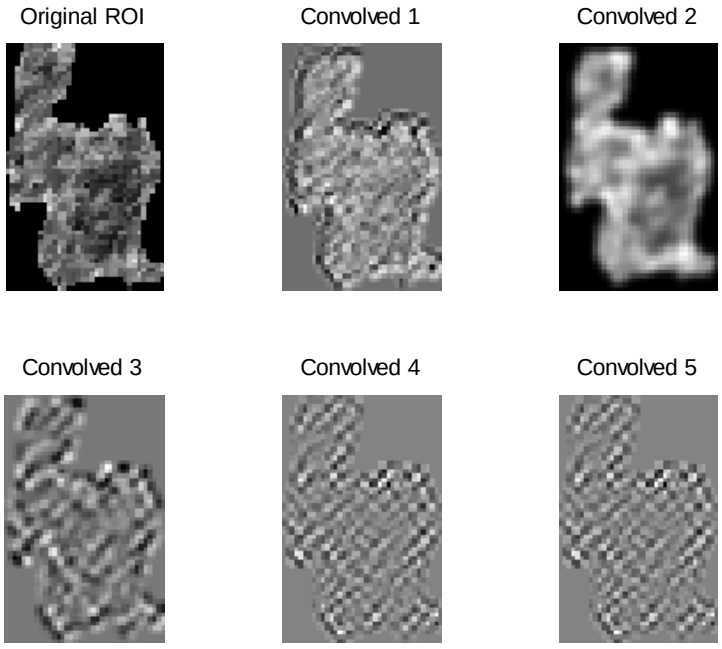


Figure 7-12 The Original ROI and the ROI after Convolution with Some of the Laws Masks

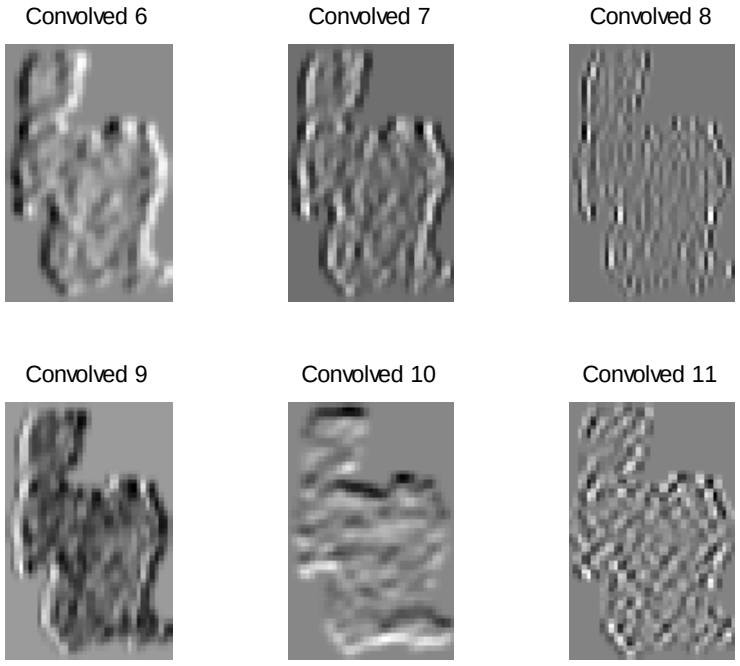


Figure 7-13 the ROI after Convolution with some of the Laws Masks

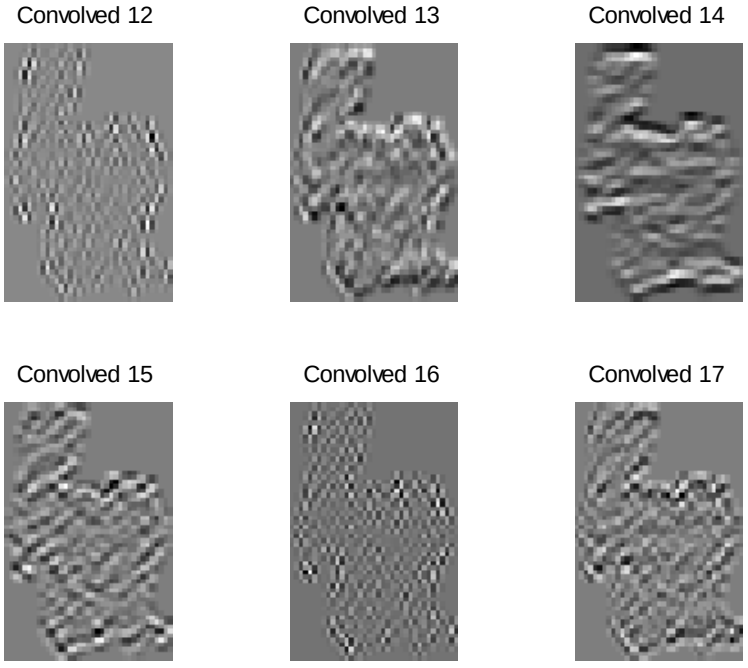


Figure 7-14 the ROI after Convolution with some of the Laws Masks

From the figures, it is clear that the resulting feature images look different which shows that each one of them carries a piece of information. The variances of the obtained feature images are then obtained to represent the energy and are considered as features for each ROI. Therefore, there are 25 different features that are constructed from each identified ROI.

7.4 Constructed Features Statistical Analysis

7.4.1 FD Feature Analysis

The FD features are mainly the FD measure obtained using three different methods; therefore, they are expected to be highly correlated. Hence, the FD features are statistically analyzed in this section and their correlations are calculated. Figure 7-15 shows the boxplot of the three generated FDs, where the boxplot method for data visualization was explained earlier in section 4.5. The boxplots were generated in this work using the three constructed FDs for both the cancerous regions (represented by 1 in the figure) and the non-cancerous regions (represented by 0 in the figure). C1 represents FD calculated using Box counting, C2 represents FD calculated using DBC, C3 represents FD calculated using PSD and the (*) represents an outlier. It is clear from the figure that each feature contributes by a portion to the ROI identification problem solution. Yet, each feature is considered important as each of them is depending on a different approximation. Figure 7-16 shows a sample of the correlation between two of the obtained FDs for the different ROIs. The figure shows high correlation as the data is hugging the straight-line fit. Moreover, the calculated correlation coefficient according to equation 5-1 was **0.97**, which is considered very high correlation. Furthermore, the MI between the FD features was also calculated as explained in section 5.4.1.3 and the average MI among these features is **1.3**, which is considered a large value compared to the values obtained earlier in chapter 4. The MI between the FD features and the class is also calculated according to the method explained in section 5.4.1.2 and it was 0.03, which is not considered a large value compared to the MI between any of the statistical features and the class that was discussed earlier and shown in Figure 5-9. These results prove that FD features by itself had some important information regarding the class, however, FD is not sufficient to represent the tissue texture. Therefore, the LEMs are used to assist in the tissue characterization process.

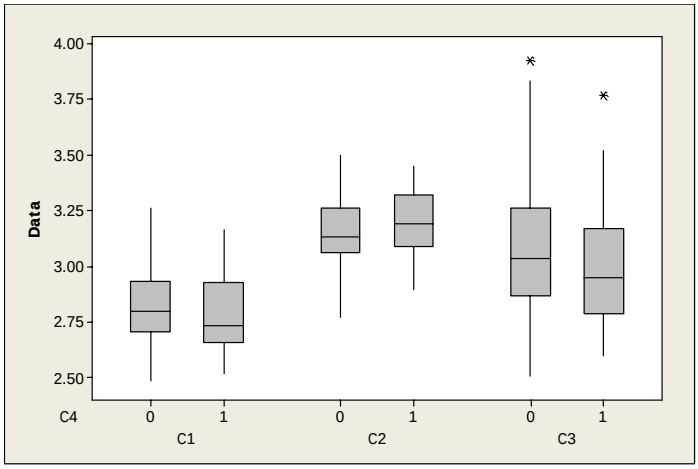


Figure 7-15 the Box plot of the fractal features

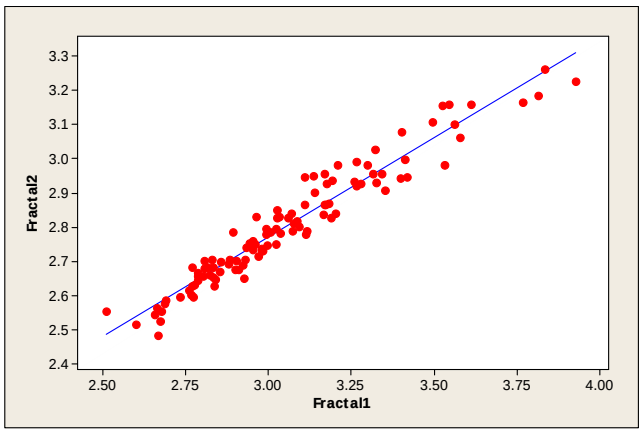


Figure 7-16 the Correlation Between two Fractal Features

7.4.2 LEMs Feature Analysis

The variances of the convolution of each ROI with each of the Laws masks are the features that constitute the LEMs feature set. Those measures are also expected to be correlated as they are all derived from the basic three masks. The correlation between these features is calculated and showed very high correlation coefficients. A sample of the correlation coefficients is shown in Table 7-1 and a sample of the figures that show this correlation is shown in Figure 7-17, Figure 7-18 and Figure 7-19. It is clear from the values shown and the figures that the LEMs are very highly correlated features. Moreover, a sample of the boxplot of three of the LEMs for the cancer and non-cancer regions is portrayed in Figure 7-20. The boxplots of three constructed features for both the cancerous regions (represented by 1 in the boxplot) and the non cancerous regions (represented by 0 in the boxplot) are illustrated where C1, C2 and C3 are the three different LEMs features. From the figure it

is clear that each feature by itself doesn't show that much difference between the cancer and the non-cancer regions, however if the information is taken from each of those features better classification accuracy can be obtained as the features are expected to complement each other. Yet, there is redundancy and correlation between some of the features

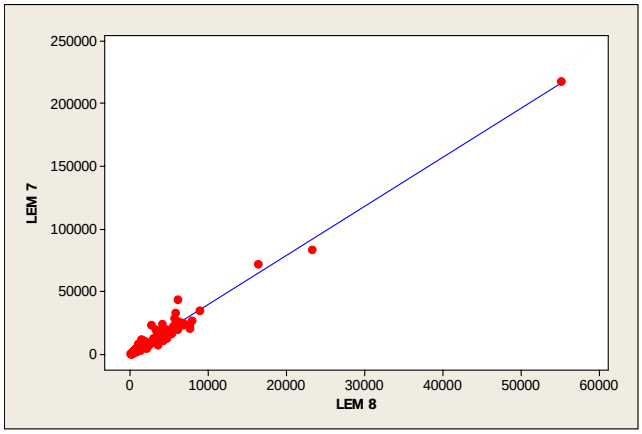


Figure 7-17 Illustration of the correlation between LEM 7 and LEM 8

Table 7-1 Correlation Coefficients among LEMS features

	LEM 2	LEM3	LEM4	LEM5	LEM6	LEM7	LEM8	LEM9	LEM10
LEM1	0.983	0.912	0.868	0.973	0.97	0.975	0.987	0.907	0.981
LEM 2	0	0.969	0.939	0.998	0.943	0.968	0.984	0.965	0.962
LEM3	0	0	0.993	0.982	0.879	0.922	0.927	0.999	0.903
LEM4	0	0	0	0.958	0.841	0.89	0.887	0.995	0.866
LEM5	0	0	0	0	0.939	0.966	0.976	0.979	0.958
LEM6	0	0	0	0	0	0.978	0.959	0.878	0.993
LEM7	0	0	0	0	0	0	0.981	0.922	0.992
LEM8	0	0	0	0	0	0	0	0.923	0.977
LEM9	0	0	0	0	0	0	0	0	0.901

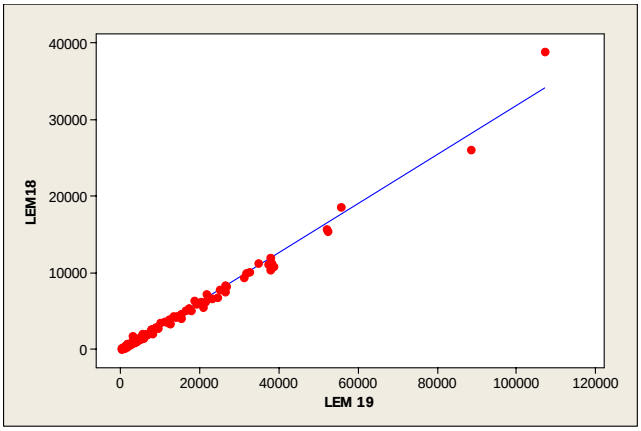


Figure 7-18 Illustration of the correlation between LEM 19 and LEM 8

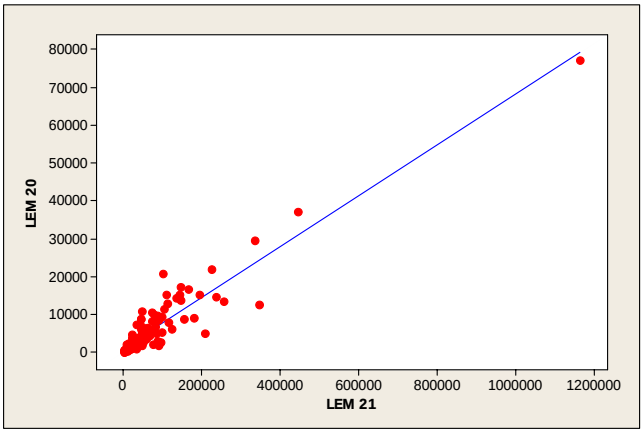


Figure 7-19 Illustration of the Correlation between LEM 20 and LEM 21

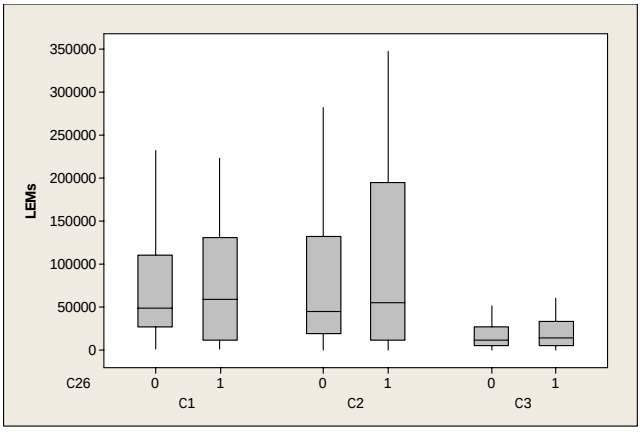


Figure 7-20 Box Plots for Three Different LEMs

In order to further analyze the relationship between the LEMs features the MI among each pair of the LEMs features is also calculated as explained in section 5.4.1.3 and a sample of the MI among some of the LEMs features (the same set that the correlation was shown) is shown in Table 7-2. It is clear from the table that the MIs among the LEMs features are much bigger than the ones obtained in Table 5-2, that corresponds to the MI among the statistical features. Having noticed the values of the correlation and the MI among the LEMs features, it could be concluded that a dimensionality reduction method should be done that will keep the information from each feature and get rid if the correlation and redundancy among features (curse of dimensionality) which will in turn obtain more accurate tissue characterization. The MI between the LEMs features and the class are also calculated as explained in section 5.4.1.2 and it was found to be **0.0087,** which is considered a small value compared to the values obtained from the statistical features. Therefore, also the LEMs need the support of the FDs to obtain good tissue characterization.

Table 7-2 MI among LEMs Features

	LEM1	LEM 2	LEM3	LEM4	LEM5	LEM6	LEM7	LEM8	LEM9	LEM10
LEM1	0	1.05	0.82	0.59	0.52	0.79	0.66	1.19	1.13	0.6
LEM 2		0	1.26	0.93	0.8	1.2	0.97	1.16	0.97	0.9
LEM3		0	0	1.6	1.3	2.5	0.71	0.99	1.33	1.5
LEM4		0	0	0	2.2	1.82	0.47	0.8	0.91	2.6
LEM5		0	0	0	0	1.49	0.57	0.7	0.82	2.3
LEM6		0	0	0	0	0	0.68	0.98	1.25	1.73
LEM7		0	0	0	0	0	0	1.16	0.2	0.63
LEM8		0	0	0	0	0	0	0	1.26	0.79
LEM9		0	0	0	0	0	0	0	0	0.92
LEM10		0	0	0	0	0	0	0	0	0

By observing the features' correlation, the features' discriminatory power and the features MI the importance for efficiently reducing the feature set is clear. Reducing the feature set with regular FS methods is not chosen in this chapter of the proposed work, as it will reject some of the features, which will degrade the classifier performance. Therefore, FE is chosen to be applied in this work to transform the constructed features into a new lower dimension domain, where all the available information is still preserved. The feature extraction and dimensionality reduction is explained in the following section.

7.5 Dimensionality Reduction

Having estimated the FD using three different methods and calculated the LEM using 25 different masks, it is expected to have redundant and correlated features, however each feature is expected ideally to have a piece of information that is missed by the other features. Therefore, traditional feature selection will not fit this feature set as it might loose some of the information that is present in the rejected features. Feature Extraction (FE) with dimensionality reduction appears to be more appealing in solving this problem. The traditional method to deal with the similar problems is the Principal Component Analysis (PCA) that is designed to model linear variations in high dimensional feature space, an approximation that is considered an over simplification to the available data set. Locally Linear Embedding systems (LLE) have recently been used for unsupervised learning and dimensionality reduction. A new feature set with lower dimension is extracted from the original feature set using each observation's neighbors' information. LLE tries to discover the non-linear structure that underlies the data under examination. Therefore, it is applied in this work for the first time in the analysis of TRUS images' features for the purpose of tissue characterization.

7.5.1 Locally Linear Embedding systems (LLE)

LLE [152] is an unsupervised learning algorithm that computes lower dimensional, neighborhood preserving embeddings of high dimensional data set. LLE attempts to discover nonlinear structure in high dimensional data by exploiting the local symmetries of local linear reconstructions. LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations do not involve local minima. Yet, it is capable of generating highly non-linear embeddings.

Each data point in the LLE algorithm and its neighbors are assumed to lie on or close to a locally linear patch of the input manifold. The LLE algorithm derives its name from:

- ***Local:*** in the sense that only neighbors contribute to each reconstruction,
- ***Linear***: in the sense that reconstructions are confined to linear subspaces.

In other words "*nearby points in the high dimensional space remain nearby and similarly co-located with respect to one another in the lower- dimensional space*."

The LLE algorithm is based on simple geometric intuitions and can be summarized as follows:

Suppose the data consist of N real-valued vectors each of dimensionality D. It is expected that each data point and its neighbors lie on or close to a locally linear patch. The local geometry of these patches is characterized by linear coefficients that reconstruct each data point from its neighbors. In the simplest formulation of LLE, one identifies nearest neighbors per data point, as measured by Euclidean distance. Reconstruction errors are then measured by the cost function:

$$\varepsilon(W) = \sum_i \left| X_i - \sum_j W_{ij} X_j \right|^2 \qquad \textbf{7-14}$$

This adds up the squared distances between all the data points and their reconstructions. The weights summarize the contribution of the j^{th} data point to i^{th} reconstruction. To compute the weights, the cost function should be minimized subject to two constraints:

- First, that each data point is reconstructed only from its neighbors;
- Second, the rows of the weight matrix sum to one

The optimal weights subject to these constraints are obtained by solving a least squares problem. In the final step of the algorithm, each high dimensional observation X is mapped to a lower dimensional vector Y representing global internal coordinates on the manifold. This is achieved by choosing d-dimensional coordinates to minimize the embedding cost function:

$$\phi(Y) = \sum_i \left| Y_i - \sum_j W_{ij} Y_j \right|^2 \qquad \textbf{7-15}$$

This cost function is based on locally linear reconstruction errors, but the weights are fixed while optimizing the coordinates Y. The LLE algorithm can be summarized as follows:

- Compute the neighbors of each data point X_i
- Compute the weights W_{ij} that best reconstruct each data point X_i from its neighbors minimizing the cost function in 7-14.
- Compute the vectors Y_i best reconstructed by the weight W_{ij} minimizing the quadratic form in 7-15.

The LLE is adopted in this work for dimensionality reduction with five neighbors and three new dimensions. The output of the algorithm is a new feature vector that is composed of two features. The original resultant feature vector, a PCA produced feature vector as well as the LLE produced feature vector are examined using SVMs classifier.

7.6 Dimensionality Reduction and Classification Results

The LLE algorithm is applied in this work where the input feature set is composed of 28 features for 108 ROIs (three FD features and 25 LEMs features). Several sizes of the output feature set were tried in this work and the best classification results were obtained using two output features. The new features are proved in this work to be uncorrelated features. The correlation between the obtained features was calculated by equation 5.1 and as found out to be zero, which proves that the LLE generates uncorrelated features. The histograms for the first LLE extracted feature for the available data set (108 ROIs) for both the cancer and non-cancer regions are shown in Figure 7-21 and Figure 7-22. The cancer regions histogram look quite different than the non-cancer regions histogram. Moreover, the statistical measures explained earlier in section 4.7.2 of those two features are calculated and those measures are shown in Table 7-3. From the table it is clear that the statistics of

the cancer regions are different from those of the non-cancer regions, which show that those extracted features are suitable for representing the tissue texture. Moreover, the boxplots that represent the LLE extracted features are shown in Figure 7-23 Figure 7-24, where the data seems to be different from the cancer and non cancer regions. This observation supports the conclusion that those features are more representative than the original feature set.

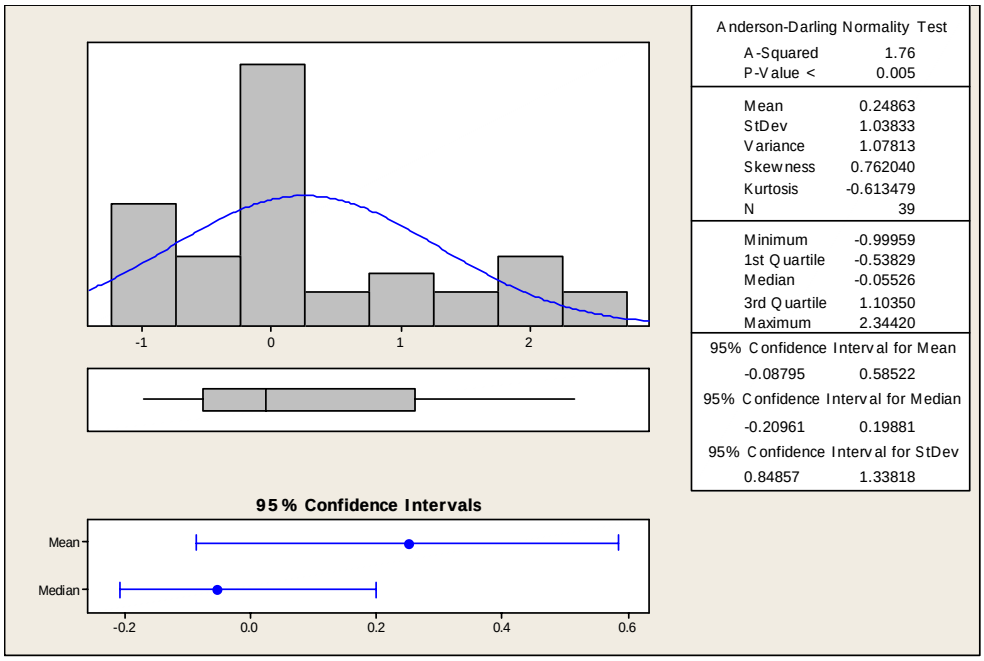


Figure 7-21 Feature 1 for cancerous regions

Table 7-3 Statistical Measures for LLE Features

		Mean	StD	Q1	Median	Q3	Skewness	Kurtosis
Cancer	Feature 1	**0.249**	**1.038**	**-0.538**	**-0.055**	**1.104**	**0.76**	**-0.61**
Non-Cancer		-0.141	0.964	-0.949	-0.356	0.354	0.97	0.2
Cancer	Feature 2	**-0.073**	**0.586**	**-0.246**	**-0.236**	**-0.006**	**3.91**	**16.34**
Non-Cancer		0.004	1.182	-0.251	-0.243	-0.139	7.59	60.33

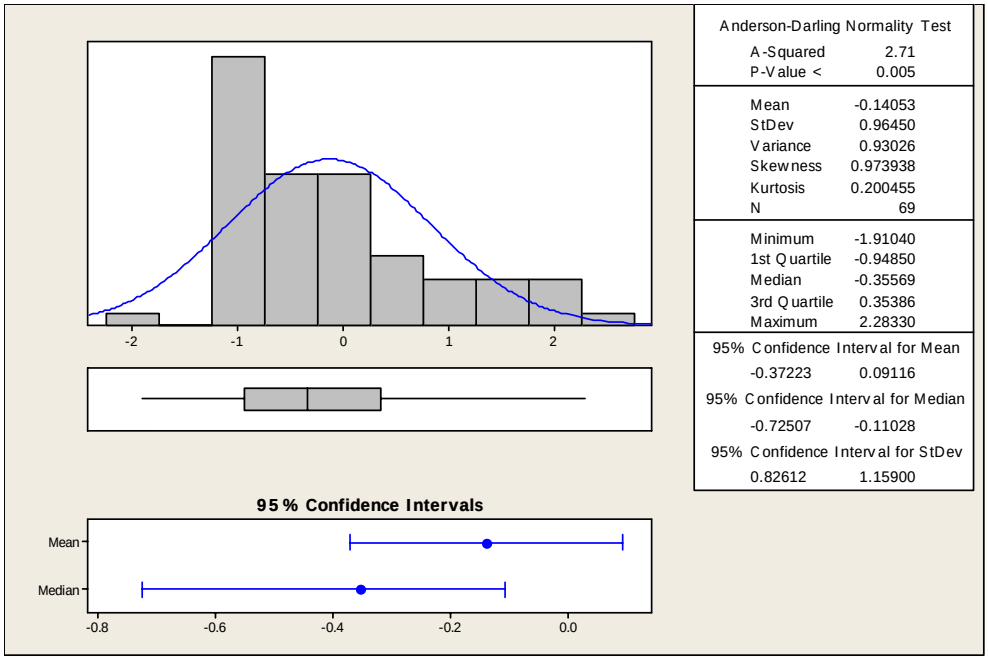


Figure 7-22 Feature 1 for non-cancerous regions

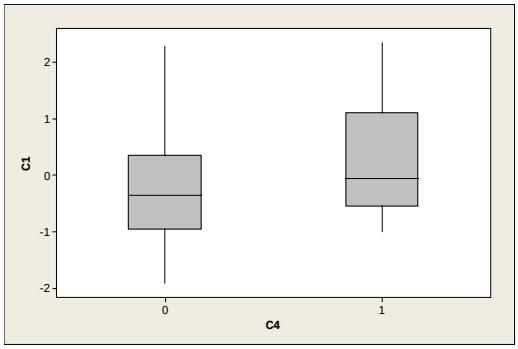


Figure 7-23 Boxplot of Feature1 for Cancer and Non-Cancer Regions

The originally obtained feature sets using the three approximations for FD estimation as well as the LEM calculated using the 25 different masks are combined and tested using the SVMs where the resulting accuracy was 66.6% with 6 misclassified cases among 18 test cases.

When applying the PCA algorithm for dimensionality reduction the accuracy did not improve with the same six misclassified cases. This can be explained by the fact that the PCA assumes linearity in the data set and uses linear transformation of the available data to the new co-ordinate system, however this assumption is considered an over simplification of the problem.

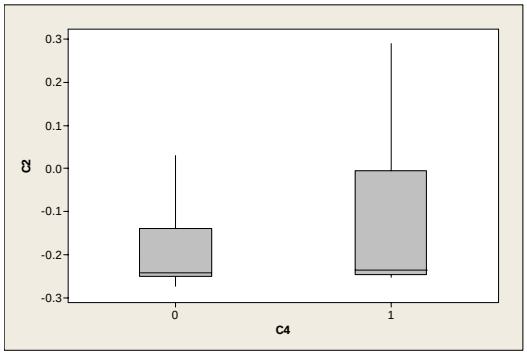


Figure 7-24 Boxplot of Feature2 for Cancer and Non-Cancer Regions

On applying the LLE algorithm to the same feature set the resulting accuracy is enhanced with 100% sensitivity, 83.3% specificity and 94.4% overall accuracy as the LLE didn't assume linearity for the whole data set but only assumed piecewise linearity, where each observation can be considered laying on a linear patch with its neighbours. The LLE also preserves the neighborhood information which keeps the important information during the transformation.

7.7 Conclusions

In this chapter the ROIs were analyzed using a feature set that is constructed using a different basis than either the statistical or the spectral feature sets that was used earlier in this thesis. The FD is obtained for the ROIs using three different approximations. The constructed FDs are statistically analyzed where high correlations and MIs are found to be present among the FD features. However, the class information that is present in each of the FD features is not big compared to those of the statistical feature. Moreover, 25 LEMs features were constructed for each ROI, the correlations and the MIs were calculated which shows the large correlation and interdependence among the LEMs features. Yet, each of the constructed FDs and LEMs features is adding a piece of information that helps in solving the problem. Therefore, to solve the tissue classification problem without loosing any of the information provided by the features, feature extraction is used. LLE is applied for FE in this chapter to obtain two features instead of the input 28 features. The obtained features are supposed to carry all the information from the original features, furthermore, the LLE extracted features proved to be uncorrelated from the statistical analysis and correlation coefficients obtained. The obtained classification results using the LLE feature set are much better than the classification results obtained using the original feature set. Using the PCA to reduce the feature space did not perform as good as the LLE because the PCA used linear transformation, which is considered an over simplification for the problem.

Chapter 8
Spectral Clustering for TRUS Image

8.1 Introduction

In the previous chapters in this thesis, CAD was used as a tool for TRUS images tissue typing where features are constructed from the echoes after the echo processing and the formation of the image. Different methods were implemented for feature construction such as statistical features that were explained in detail in chapter 4; transform based features that were explained in detail in chapter 6 and some extra texture features such as the LEMs and the FDs that were explained in chapter 7. Statistical approaches represent texture with the non-deterministic properties of the relationships between the grey levels of the image. Some other approaches represent texture on the basis of the spectral properties of the image. Moreover, there are more sophisticated textural features that are determined by Fractal models that relate local features to a specific global constraint which is the fractal dimension. The proposed framework so far obtained excellent tissue characterization techniques that did obtain cancerous regions that were not identified by the radiologist (our only available gold standard). However, the radiologist did confirm that he might have missed some of those regions. Therefore, there is a need for a technique that can support the decision of the implemented feature based algorithms' obtained tissue classification results.

The CAD system proposed in this Chapter has a different foundation than the previously proposed systems in this thesis as the earlier algorithms rely mainly on image processing and feature analysis techniques. The TRUS image segmentation is achieved in this chapter using the Spectral Clustering (a recently proposed data clustering method). In the case of image segmentation, Spectral Clustering treats the image as a weighted undirected graph and finds the segmentation result by obtaining the minimum cut of this weighted graph based on the graph theory methods.

The spectral clustering method is not yet considered a well established ultrasound image segmentation method. Therefore, there is a need for testing the credibility of this algorithm in dealing with ultrasound images, hence justifying its application for the prostate image ROI identification problem (prostate region segmentation). The prostate boundary detection is considered a problem whose solution is well defined, therefore this problem is solved in this chapter using the spectral clustering algorithm and the obtained solution is compared to the available gold standards. Hence, the problem that is first tackled in this chapter to prove the credibility of spectral clustering is prostate boundary detection. The resulting boundaries are then compared to those highlighted by the expert radiologist. The results of the prostate boundary detection is shown in section 8.4.1.

The Gabor filters followed by extensive feature analysis were used earlier in this thesis to identify the cancerous ROIs. However, the proposed algorithm in this chapter is faster on the expense of giving a rough estimate of the cancerous regions than the earlier presented work. The algorithm proposed in this chapter can then be used for suspicious regions estimation in an online application. It can also be

used to support the decisions obtained using the feature analysis especially if the later contradicts the radiologist's decision. The use of spectral clustering for ROI identification is shown in section 8.4.2.

8.2 Spectral Clustering

As discussed earlier in chapter 3, the Human Visual System (HVS) can effectively identify objects in a scene and can often segment the scene to coherent segments or clusters. There has been a tremendous amount of research done to achieve the same level of performance obtained by the HVS. Various methods have been introduced in literature to segment ultrasound images such as Kalman filters [153] and statistical shape models [154] in which the prostate is segmented from the background in order to determine the gland volume. However most of these methods require large amount of human interaction. The Gabor filters were used in this thesis as a method to internally segment the prostate gland that was already manually segmented from the background. A novel method utilizing the Gabor filter was implemented, explained and applied for the TRUS images segmentation in Chapter 3. The ROIs obtained in Chapter 3 did segment regions that were not identified by the radiologist, whose ceiling in cancer recognition is not more than 66 %. The radiologist confirmed that the extra regions obtained by the ROI identification algorithm proposed in this thesis are suspicious. The obtained ROIs were further analyzed using the texture features that were explained thoroughly in the previous chapters. Spectral Clustering methods are applied in this chapter in order to support the decision obtained in the previous chapters whenever needed. Spectral Clustering is chosen to achieve this task due to the fact that its segmentation results are obtained very fast (12 seconds average) and it doesn't depend on any of the features that were obtained earlier in the previous Chapters. Spectral Clustering have been introduced for data clustering and were applied in different fields. The spectral clustering usually represents the data by a weighted graph and the eigenvectors of the affinity (or similarity) matrix are used for the segmentation [155]. In the problem of image segmentation the image pixels are considered as the data points as shown in [156].

8.2.1 Graph Based Image Segmentation

Given an image I, a graph $G = (V, E, W)$ is constructed with the pixels represented by the graph nodes V , and the pixels within a distance $V \le G_r$ are connected by a graph edge E . The weight $W(i, j)$ measures the likelihood of pixel i and j being in the same cluster. Partitioning of this graph represents the image segmentation [156, 157, 157].

8.2.2 Assigning Weights to Graph Edges

The pair-wise pixel affinity graph determines the segmentation accuracy. Therefore as recommended in [157] two simple local grouping cues are used which are the intensity and contours.

- **Grey Level Intensity**: neighboring pixels with close intensity are most likely to be in the same region.

$$W_i(i,j) = e^{-\frac{\left\|X_i - X_j\right\|^2}{\sigma_x} - \frac{\left\|I_i - I_j\right\|}{\sigma_I}} \quad \textbf{8-1}$$

Where X_i and I_i represent pixel location and intensity.

Connecting pixels considering only intensity and location usually gives bad segmentation due to the texture that is present in the TRUS images. Therefore the principal image contours (edges) are also considered for the segmentation of TRUS images.

- ***Dominant Contours***: the image edges are considered useful when the neighboring regions have the same cutter. The affinity between two pixels is calculated by measuring the image edges between them.

$$W_C(i,j) = e^{-\max_{x \in line(i,j)} \left\|Edge(x)\right\|^2 / \sigma_c} \quad \textbf{8-2}$$

Where $line(i,j)$ is a straight line joining pixels i and j and $Edge(x)$ is the edge strength at location (x).

The two cues are combined in this work in the form:

$$\sqrt{W_I(i,j) \times W_C(i,j)} + \alpha W_C(i,j) \quad \textbf{8-3}$$

8.2.3 Spectral Clustering Segmentation Algorithms

In [157], a clustering algorithm based on thresholding the largest eigenvector of the affinity matrix was suggested. While in [159] the authors have argued for using a totally different eigenvector for solving these types of segmentation problems. Rather than examining the first eigenvector of W they examined the generalized eigenvectors. Let D be the degree matrix of W :

$$D(i,i) = \sum_j W(i,j) \quad \textbf{8-4}$$

The generalized eigenvector y_i is a solution to:

$$(D - W)y_i = \lambda_i D y_i \quad \textbf{8-5}$$

Solving the generalized eigenvector minimizes the Ncut which in turn produce the optimum segmentation as proved in [156]. In this case they used the generalized eigenvector corresponding to the second smallest eigenvalue. They suggested thresholding this second eigenvector to obtain the

segmentation result. This method is adopted for the application of TRUS image segmentation and it yields to a segmentation that minimizes the normalized cut:

$$Ncut(A,B) = \frac{cut(A,B)}{asso(A,V)} + \frac{cut(A,B)}{asso(B,V)} \quad \textbf{8-6}$$

where: $A \cup B = V$ *and* $A \cap B = 0$, $cut(A,B) = \sum_{i \in A, j \in B} W(i,j)$ and $asso(A,V) = \sum_{j} \sum_{i \in A} W(i,j)$

Therefore, the solution to the segmentation problem minimizes the affinity between groups normalized by the affinity within the same group. In this work, the spectral clustering is used for the first time for TRUS image segmentation using the approach proposed in [159].

8.3 Proposed Algorithm Implementation

The segmentation proposed algorithm is composed of the following steps:

1. **The edge map** of the TRUS image is obtained using Canny edge detection method
2. **The Affinity matrix** is created using equation 3.
3. **The eigenvectors** are calculated and reshaped to be shown in the figures.
4. **Eigen vector Discretization**: In the best case scenario the second smallest eigenvector should take on two discrete values and the signs of the values can tell how to partition the graph. The second smallest eigenvector obtained in our case is a continuous vector; therefore, it needs to be discretized in order to obtain a splitting point to obtain the clustering. In this work the splitting point that minimizes Ncut equation (8.6) is chosen. The algorithm applied in this work is similar to that proposed in [156].

8.4 Experimental Results

In this section, some TRUS images' segmentation results are presented. First the prostate was segmented from the background (boundary detection) and the result segmentation is compared to the radiologist segmentation. The Eigen vectors of the affinity matrix are also shown and the relation between those Eigen vectors and the desired segmentation is examined. The obtained segmentation was then evaluated using the overlap area measure and was also examined visually. The obtained boundaries were very close to the ones obtained by the expert radiologist. The results of this work are shown in section 8.4.1. Therefore, from the first section, the spectral clustering proved its ability to deal segment the ultrasound images accurately. Hence its segmentation results should be trusted in the case of prostate region segmentation. Therefore, the segmented prostate is then further segmented where the gland is subdivided into different regions. The obtained regions were found to be correlated to the regions marked by the radiologist and also with the regions obtained earlier in this thesis using the Gabor filter. Moreover, the relationship between the obtained ROI segmentation and the eigenvectors of the affinity matrix was examined . The ROI identification is shown in section 8.4.2.

8.4.1 Prostate Boundary Segmentation

Spectral Clustering was applied for medical image segmentation only in one recent publication [160] where it was applied mainly to artificial ultrasound images and was tested on a couple of real ultrasound images obtained from vivo. Therefore more justification needs to be done to accept the algorithm segmentation especially in the case of low ground truth accuracy. In order to test the validity of applying the Spectral Clustering for the TRUS image segmentation for the purpose of tissue classification and to ensure that the internal segmentation of the gland is acceptable, the algorithm is tested in this work with a problem whose solution is well defined. Therefore, Spectral Clustering is utilized to segment the prostate gland from the background (boundary detection) and the results are compared to those obtained by the radiologist (considered better ground truth than recognizing the cancerous regions). Several Prostate images (29 TRUS images) are segmented from the background using the proposed Spectral Clustering algorithm with high accuracy. The results are shown in the following figures. Figure 8-1 shows the original image, the radiologist's map and the second eigenvector reshaped to an image. Figure 8-2 shows all the obtained affinity matrix eigenvectors. Figure 8-3 shows a) obtained segmentation using the implemented spectral clustering algorithm, b) the contour map obtained using the spectral clustering algorithm overlapped on the original TRUS image. Figure 8-4 shows a) the manually segmented prostate, b) the Spectral Clustering segmented prostate and c) the two prostate masks overlapped. From the figures it is clear that the proposed spectral clustering algorithm obtained high accuracy in segmenting the prostate from the background. Moreover, the second eigenvector is the most correlated one with the desired segmentation. The following figures from Figure 8-5 to Figure 8-31 show ten more segmented prostate images that have various shapes, orientations, locations and grey levels. The figures proved the ability of the spectral clustering algorithm to segment the prostate accurately without any prior knowledge about the gland. Moreover, the spectral clustering segmented prostate is compared to the manually segmented prostate using the overlap area measure. The overlap area measure is calculated by dividing the common area between the manually segmented prostate and the spectral clustering segmented prostate by the area of the manually segmented prostate. A sample of these accuracy results are shown in Table 8-1 Prostate Boundary Detection Accuracy and the average percentage overlap area for all the tested images was found to be 94% which proves the superiority of this method in obtaining the prostate boundary accurately regardless of its position, orientation and grey level. The proposed algorithm is considered a merit in the boundary segmentation, as it didn't need any seed points or preliminary contours to be defined. Therefore it is expected that it will produce good results for the cancer region segmentation which is discussed in the following section.

Table 8-1 Prostate Boundary Detection Accuracy

Image #	1	2	3	4	5	6	7	8	9	10	11	12
% Acc.	93	93.04	93.02	93.47	93.03	93	92.98	95	94.48	98	93.76	95

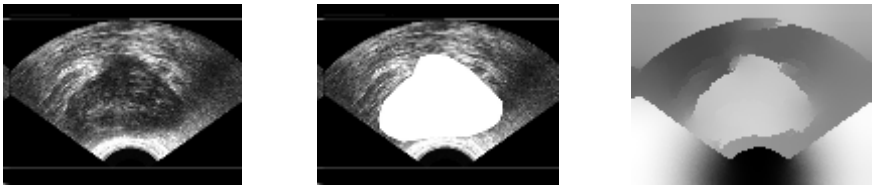


Figure 8-1 The Original TRUS Image1, Manually Segmented, Second Eigenvector

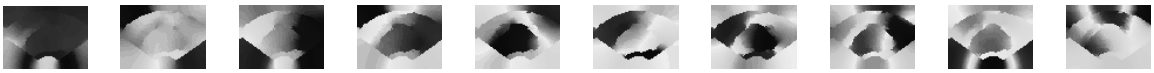


Figure 8-2 All Obtained Eigenvectors

a b

Figure 8-3 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-4 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

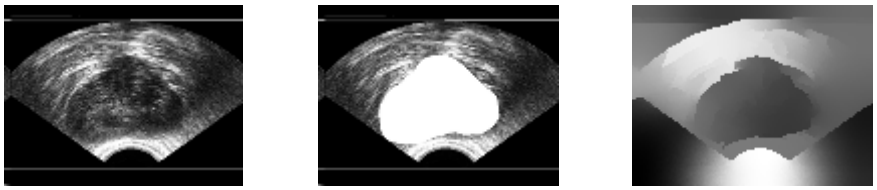


Figure 8-5 The Original TRUS Image2, Manually Segmented, Second Eigenvector

a b

Figure 8-6 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-7 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

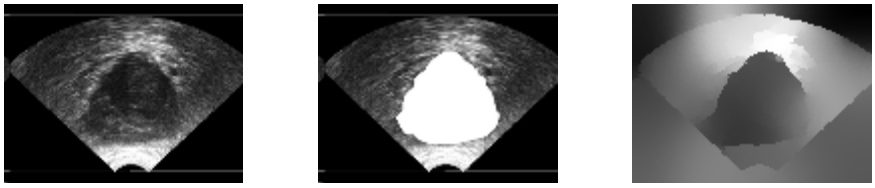


Figure 8-8 Original TRUS Image3, Manually Segmented, Second Eigenvector

a b

Figure 8-9 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-10 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap image

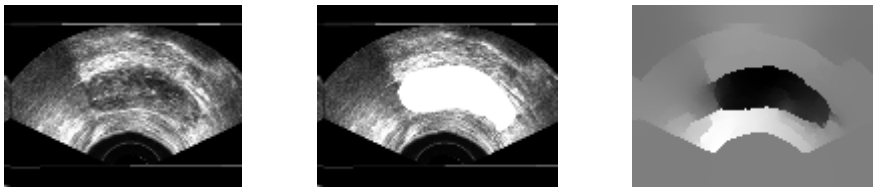


Figure 8-11 Original TRUS Image 4, Manually Segmented, Second Eigenvector

a b

Figure 8-12 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-13 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

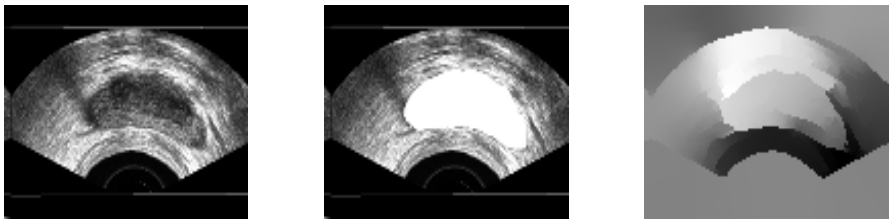
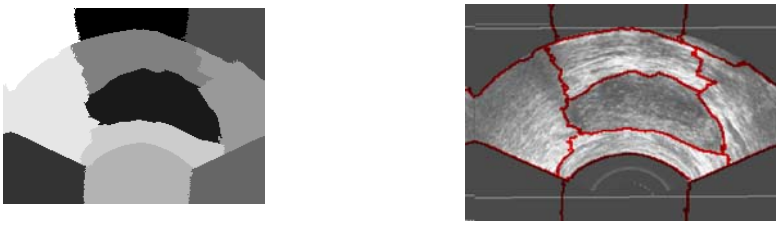


Figure 8-14 Original TRUS Image 5, Manually Segmented, Second Eigenvector



a b

Figure 8-15 a) Segmentation Result, b) Obtained Contours

Figure 8-16 a) Manual Segmentation, b) Spectral Clustering Segmentation, c)Overlap Image

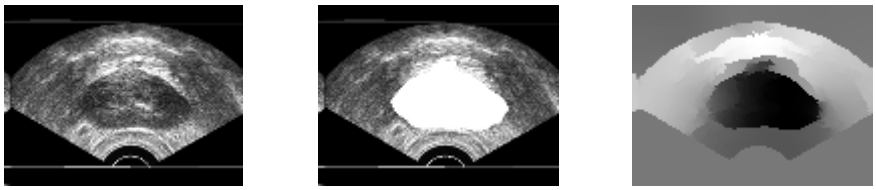


Figure 8-17 Original TRUS Image 6, Manually Segmented, Second Eigenvector

a b

Figure 8-18 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-19 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

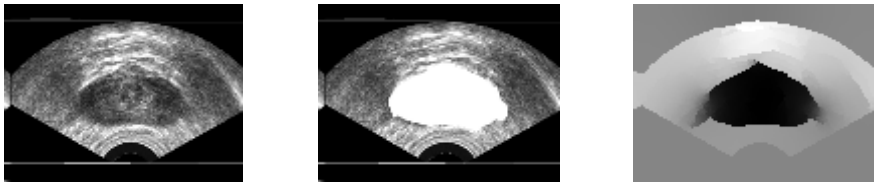


Figure 8-20 Original TRUS Image 7, Manually Segmented, Second Eigenvector

a b

Figure 8-21 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-22 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

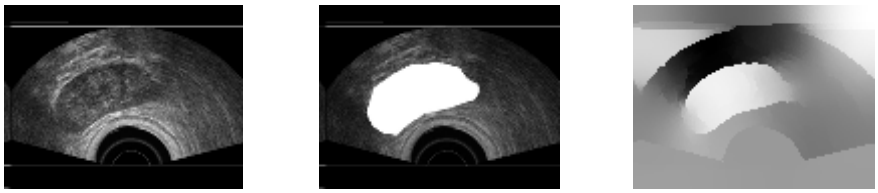


Figure 8-23 Original TRUS Image 8, Manually Segmented, Second Eigenvector

a b

Figure 8-24 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-25 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

Figure 8-26 Original TRUS Image 9, Manually Segmented, Second Eigenvector

a b

Figure 8-27 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-28 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

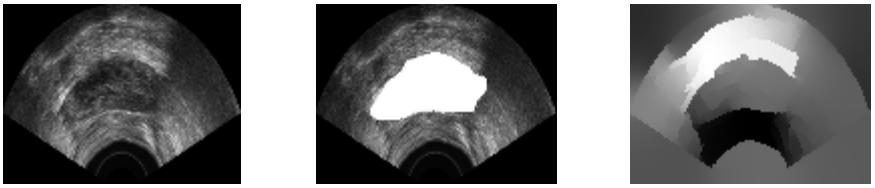


Figure 8-29 Original TRUS Image 10, Manually Segmented, Second Eigenvector

a b

Figure 8-30 a) Segmentation Result, b) Obtained Contours

a b c

Figure 8-31 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Overlap Image

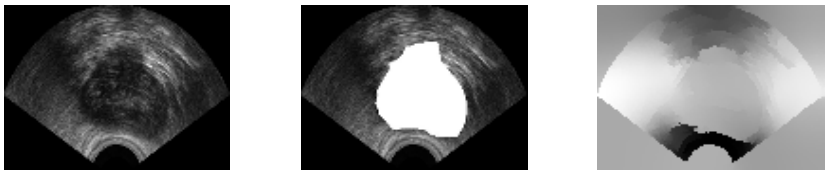


Figure 8-32 Original TRUS Image 11, Manually Segmented, Second Eigenvector

a b

Figure 8-33 a) Segmentation Result, b) Obtained Contours

Figure 8-34 a) Manual Segmentation, b) Spectral Clustering Segmentation, c) Difference

8.4.2 Identifying Cancer regions

The spectral clustering image segmentation algorithm is mainly used in this work for segmenting cancerous regions. However it was applied in the previous subsection for the prostate gland segmentation from the TRUS image and it obtained high segmentation accuracy which proves its capability of dealing with TRUS images effectively. Therefore, it is used in this sub-section for cancer recognition from the manually segmented prostate.

The Gabor filters followed by extensive feature analysis were used earlier in this thesis to identify the cancerous ROIs. However, the proposed algorithm in this chapter is faster on the expense of giving a rough estimate of the cancerous regions than the earlier presented work. The algorithm proposed in this chapter can then be used for suspicious regions estimation in an online application. It can also be used to support the decisions obtained using the feature analysis especially if the later contradicts the radiologist's decision.

8.4.2.1 Typical prostate shape

The proposed spectral algorithm was successful for capturing the suspicious cancerous regions from the TRUS images with typical prostate shape. For the typical gland shape shown in from Figure 8-35 the original image and the corresponding doctor's prostate region segmentation are shown, Figure 8-36 shows all the obtained eigen vectors, Figure 8-37 shows the second smallest eigen vector that shows very high correlation with the desired segmentation, Figure 8-38 shows the obtained segmentation and the corresponding regions contours, while Figure 8-39 shows the overlap of the obtained segmentation and the manual segmentation. Another example is shown in the following figures from Figure 8-40 to Figure 8-43**.** The results show that the algorithm was successful in identifying the doctor's suspicious regions in the typical gland. Moreover, it is also able to segment regions that were recognized by the Gabor filter and missed by the radiologist as shown in Figure 8-42 and comparing this segmentation result to those obtained by the Gabor filter that was shown in Figure 3-21. The proposed Spectral Clustering algorithm did capture the information that was hidden from the radiologist's decision and can't be seen by the naked eye which confirms and supports the decision of the recognition methods explained earlier in the thesis. Therefore, it can be concluded that the Spectral Clustering algorithm segmented regions that were missed by the radiologist, yet still carry the same information of the radiologist identified regions which proves the superiority of the proposed algorithm.

8.4.2.2 Unusual Prostate Shape

The spectral clustering algorithm also shows excellent segmentation even if the prostate shape is not the typical shape. The ability of the proposed algorithm is clear from the figure sets from Figure 8-44 to Figure 8-51. It is clear from the figures that the regions that are marked by the doctor can be predictable by the second smallest eigenvector and the segmentation algorithm did capture a rough estimate of the suspicious regions and their location in the gland. Moreover, there is a great consistency between the regions obtained by the Gabor filter presented in Figure 3-23 and Figure 3-26 and the segmented regions by the Spectral Clustering method. This result proves the ability of the proposed algorithm even for the atypical prostate shape a case that is usually confusing the registration methods. This proves the ability of the algorithm to be used as a preliminary online estimate for the cancerous regions as well as a support for the previously explained methods.

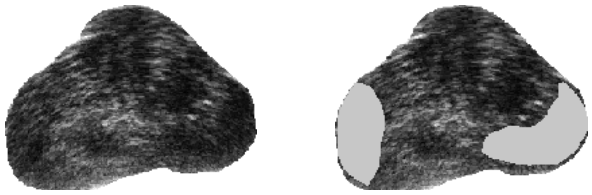


Figure 8-35 Prostate TRUS Image 1 with the desired internal segmentation

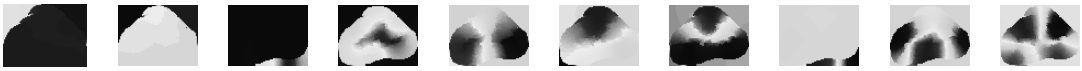


Figure 8-36 Obtained Eigen-vectors where the second smallest shows correlation with the ROI

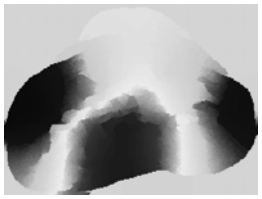


Figure 8-37 The Second Eigen Vector enlarged

a b

Figure 8-38 a) Segmentation Result, b) Obtained Contours

a b

Figure 8-39 The Overlap between each of the Identified regions and the Manual Segmentation

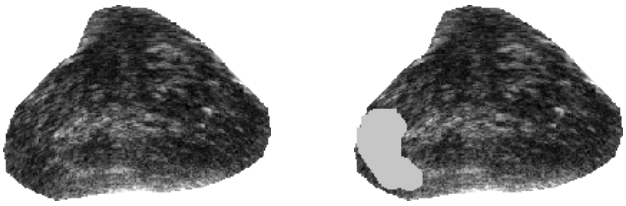


Figure 8-40 Prostate TRUS Image 2 with the desired segmentation

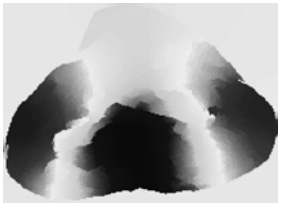


Figure 8-41 The Second Eigen Vector enlarged

a b

Figure 8-42 a) Segmentation Result, b) Obtained Contours

Figure 8-43 The Overlap between the Identified region and the Manual Segmentation

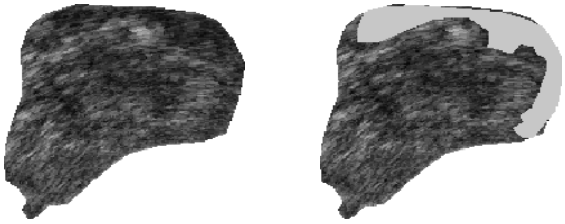


Figure 8-44 Prostate TRUS Image 3 with the Desired Segmentation

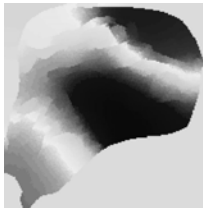


Figure 8-45 The Second Eigen Vector enlarged

a b

Figure 8-46 a) Segmentation Result, b) Obtained Contours

Figure 8-47 The Overlap between the Identified region and the Manual Segmentation

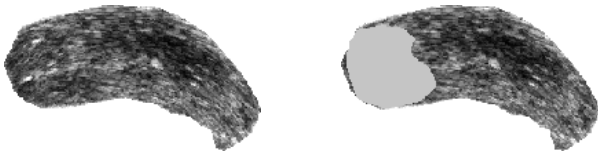


Figure 8-48 Prostate TRUS Image 4 with the Desired Segmentation

Figure 8-49 The Second Eigen Vector enlarged

a b

Figure 8-50 a) Segmentation Result, b) Obtained Contours

Figure 8-51 The Difference between the Identified region and the Manual Segmentation

8.4.2.3 Recognizing relatively big cancerous regions

It is clear from the figure sets shown from Figure 8-52 to Figure 8-63, that the algorithm is capable of recognizing relatively big cancerous regions regardless of either the prostate shape or the spatial location. Moreover, by checking the regions obtained by the Gabor filter presented in Figure 3-28 and Figure 3-29, it can be observed that there is agreement between the two methods in identifying the suspicious regions. Therefore, Spectral Clustering can be used effectively as a decision support tool for radiologists for manual region segmentation during the imaging process.

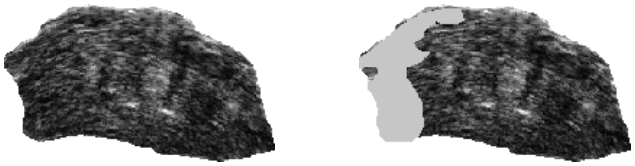


Figure 8-52 Prostate TRUS Image 5 with the Desired Segmentation

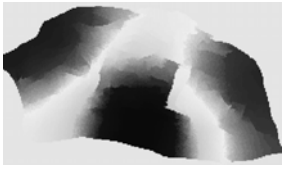
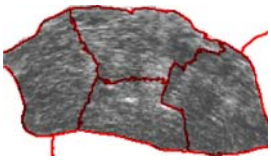


Figure 8-53 The Second Eigen Vector enlarged

a



b

Figure 8-54 a) Segmentation Result, b) Obtained Contours

Figure 8-55 The Difference between the Identified region and the Manual Segmentation

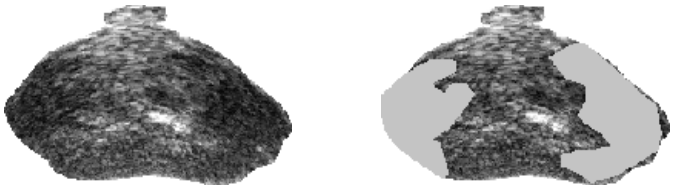


Figure 8-56 Prostate TRUS Image 6 with the Desired Segmentation

Figure 8-57 The Second Eigen Vector enlarged

a b

Figure 8-58 a) Segmentation Result, b) Obtained Contours

a b

Figure 8-59 The Difference between each of the Identified regions and the Manual Segmentation

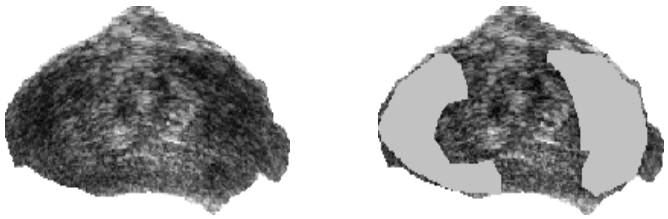


Figure 8-60 Prostate TRUS Image 7 with the Desired Segmentation

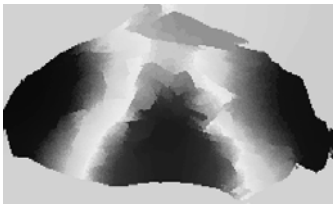


Figure 8-61 The Second Eigen Vector enlarged

a b

Figure 8-62 a) Segmentation Result, b) Obtained Contours

a b

Figure 8-63 The Difference between each of the Identified regions and the Manual Segmentation

From the above results it is clear that the spectral clustering can be used as a preliminary estimate for the cancerous regions during the imaging procedure. Moreover it can be used as a support for the decision making either by the radiologist or by the other CAD systems. It can also estimate the prostate volume, location and size from the TRUS image effectively. The second smallest eigenvector

proved to be consistent with the radiologist manual segmentation for the two different segmentation problems tackled in this chapter.

8.5 Conclusions

In this chapter the recently proposed spectral clustering method is adopted and applied for the first time for the purpose of TRUS image segmentation. Spectral Clustering is able to segment the prostate from the background effectively and accurately without any preliminary estimate of either a contour or a seed point. Moreover, the spectral clustering method is able to roughly segment the cancerous regions that proved consistency with the regions identified by the doctor and by the Gabor filter. The method did obtain good results for typical as well as atypical prostate shapes. It could also recognize relatively big cancer regions regardless of the prostate shape and the spatial location of the cancer within the gland.

The algorithm obtained fast excellent estimates that can give rough prostate volume as well as cancerous regions segmentation which can be used for online application. However in order to obtain more accurate tissue characterization the different texture features combined with the Gabor filter ROI segmentation method should be used.

Chapter 9
Conclusions and Contributions

9.1 Summary

The CAD cancer recognition systems utilizing TRUS images depend mainly on the amount of information extracted from the TRUS images. This information is much more than just the grey level values at each pixel. The recognition accuracy depends mainly on the quality of the extracted information (features) where bad features result in bad recognition accuracy. The available extraction methods using TRUS rely on either the TRUS images or the RF signals. The use of TRUS images is more appealing and is more widely used, as this kind of data is much easier for the researchers to obtain and visualize. However, the RF signals produced good recognition accuracies. There are several deficiencies in the present CAD cancer recognition using TRUS images.

- First, most of the image-based algorithms presume that the co-occurrence features are the only information (feature set) present in the TRUS image. These methods ignore some of the information that is present in the TRUS image (less information).
- Second, the constructed raw features are used directly for classification without performing any analysis or evaluation for the quality of the features. In this case, bad or redundant features might be used resulting in classifier confusion and bad accuracy (more information).
- Third, in their analysis, methods might utilize the whole image (start from vacuum) which might lead to dividing the image into small squares without any prior image scanning. These squares might have mixed cancer and non-cancer tissues which may lead to bad misleading features.
- The existing methods might also use the regions marked by the radiologist whose accuracy is limited to approximately 65%. This approach makes the human visual system the ceiling of our seach, where we cannot go beyond it.
- Fourth, in the methods using the RF signal, excellent classification results are obtained, however, the data is very hard to collect, store, manipulate and visualize.

In this thesis, several new methods are developed for improving the deficiencies that exist in the available cancer recognition system. The proposed TRUS image analysis have the following advantages:

1. The proposed algorithm introduces novel feature construction methods for TRUS images.
2. The novel feature sets are analyzed and further reduced by proposing new feature selection algorithms as well as a new feature extraction algorithm.

3. Moreover, the proposed algorithm introduces the idea of studying only the highly suspicious regions in the gland instead of studying the whole gland, which makes the analysis simpler and more precise.
4. The proposed algorithm generalizes the ideas used for RF signals to be applied to the TRUS images by scanning the TRUS images to form signals.
5. The proposed techniques are based on different foundations such as image processing, signal processing, pattern recognition and graph theory. This is important where different methods support the decision of each other.

When it comes to choosing the suspicious regions, the Gabor filter is used due to its high localization property in both the frequency domain and the spatial domain. The ROI identification proposed in chapter three is considered a smart way to select the suspicious regions for further analysis. It depends on both the frequency information and the spatial information. Moreover, the expert knowledge was also integrated into the system to obtain the smart ROI identification algorithm. The proposed ROI identification algorithm was validated and its capabilities were demonstrated by applying it to the available TRUS images and it was able to identify most of the regions that were identified by the expert radiologist. It was also capable of identifying regions that were missed by the radiologist while these regions attain the same characteristics as the radiologist identified ones.

The existing feature construction techniques using TRUS images depend mainly on the co-occurrence matrix features, which produced acceptable results but considered an expensive method. Moreover, it relies mainly on the second order statistics, which is the cap of the HVS. The proposed feature construction methods in this work using TRUS images are three manifold.

- First, a new statistical feature set that is composed of second and higher order feature sets were used in chapter four. These statistical features obtained excellent classification accuracy. The second order statistics used are the GLDM (co-occurrence) features as well as the GLDV features which are considered less expensive than the GLDM, yet it produces excellent classification results.
- Second the TRUS images' ROIs were used in chapter 6 to obtain the ROIs' signals that were introduced and proved an excellent way in representing the TRUS ROIs. Chapter 6 introduced a new method of visualizing the ROIs in the signal form which allows for the use of the RF recognition techniques to the images' ROIs. Moreover, new features were constructed from those signals based on the signal spectral analysis. The obtained spectral features provided excellent classification accuracy especially the ESPRIT features.
- Third, a novel feature set composed of the fractal dimension and the LEMs were constructed for each of the identified ROIs in chapter seven. Furthermore, those features were evaluated and found out to attain excellent recognition power.

Feature analysis is considered a vital stage in electing the best feature subset in terms of cancer recognition. Unfortunately, feature analysis was ignored by researchers in the field of prostate tissue classification from TRUS images. They used all the features they constructed without performing any

feature analysis. Feature analysis was introduced in this work in chapters 4, 5, 6 and 7 where the constructed features' histograms are shown and their statistical measures are calculated. Moreover, the correlation among the different features was also investigated which leads to the conclusion that there is some redundant and correlated information present in the constructed feature sets.

Therefore, Different FS methods were adopted in this work to get rid of this redundant and correlated information to obtain precise classification. FS is categorized into two main groups, which are namely classifier dependent and classifier independent. The first FS avenue introduced in this work is classifier independent where a feature subset was chosen depending on the features information content regardless of the classifier. The MIFS was adopted in chapter five as a form of classifier independent feature selection where a feature subset was chosen that obtained excellent classification results.

Furthermore, in chapter five another avenue for FS is introduced which is the classifier dependent FS. The classifier dependent FS methods adopted in this work are based on Alife optimization algorithms that try to mimic the behavior of different creatures. This is the first time to the best of the author's knowledge that optimization algorithms are used for TRUS images feature selection. The adopted methods namely the PSO and the ACO do not suffer from the nesting problem that is present in the existing feature selection methods. Moreover, they do not require the same computational cost required for the exhaustive search techniques. Yet, they obtained an optimum feature subset that obtained the best possible accuracy result considering the available gold standards. From chapter five the Alife techniques were recommended over the MIFS as they obtained the same classification accuracy with less computational effort and parameter selection.

One of the major problems that were present in this research is the lack of the accurate gold standards as the available gold standards for this work was the radiologists' stained images which has a maximum accuracy of 65%. Therefore, in order to support our decision, another method should be utilized. Spectral clustering is introduced in this work. Spectral Clustering does not depend on extracting features and classifying the regions accordingly. On the contrary, it depends on converting the image to a weighted graph, which is partitioned using the graph theory techniques. The segmentation results obtained by the spectral clustering algorithm supported undoubtly the decisions obtained in this thesis.

9.2 Contributions

The contribution of the work in this thesis originates from the capability it provides to construct different precise automated robust cancer recognition avenues based on TRUS images' feature analysis. The proposed algorithms can detect the suspicious regions, construct a full library of informative features, analyze the constructed features, classify the different regions and finally support the ROI classification using a different foundation algorithm. The five tasks incorporated in such a system were thoroughly investigated; then a set of algorithms that outperform their counterparts in literature, to the best of the author's knowledge, to accomplish these tasks, were proposed. This is the first time that those five tasks are tackled and integrated in one analysis. The

previous work usually focused on one of those presented tasks. The following points summarize the contributions of this research:

1. The ROI identification algorithm is proposed to eliminate the need to study the whole image and to ensure adequate image segmentation for cancer recognition. The algorithm proved to be accurate, precise and consistent.
2. TRUS feature construction methods were thoroughly studied. Most of the proposed methods in literature either require the RF signals and/or use only the second order statistical features (GLDM). The proposed higher order statistical features, spectral based features, model based features are proved to be informative enough for cancer recognition and can precisely identify cancerous regions.
3. The existing methods usually used the constructed features directly which results in the existence of the curse of dimensionality. The proposed feature selection and dimensionality reduction techniques allow for the accurate selection of feature subsets that guarantee the best classification accuracies. The proposed methods take into account, the feature information contents, the accuracy results and the transformation to new co-ordinate system.
4. Three different classifiers were compared in this work and the SVMs classifier proved to possess the best recognition ability due to its ability to deal with non-linearly separable data sets.
5. Scanning the ROIs to form ROI signal was proposed for the first time for the case of TRUS images. This approach is capable of utlizing the methods that were originally applicable for RF signals to be used for TRUS images feature analysis and tissue classification.
6. Different spectral features were introduced for TRUS images feature analysis and those features were compared regarding the tissue texture. It was found out that the wedge and ring filters are not adequate for the application because of the irregular ROI shape. It was also found that the modified periodogram was not adequate as the data set is short which leads to spectral leakage. Furthermore, combining all the constructed spectral features obtained excellent classification accuracy which proves that each feature set has its own piece of information that is important for the classification.
7. Moreover, It was found out that in the case of spectral features, the classification results that are obtained using only the ESPRIT feature set are as good as those obtained using all the estimated PSD estimation methods. This leads to a conclusion that the ESPRIT features are more adequate and robust to this application, as it did not require the construction of any more features, which reduces the computational cost and increases the accuracy of the proposed recognition algorithm, as there is no need to construct more features while only one feature set obtains the same results.
8. The identified regions were classified using the different proposed feature subsets and the classification results were consistent, which makes the decision obtained using this system

well supported even in the absence of gold standards. This proves the proposed systems' ability to classify the image tissues when the pathology identified gold standards are made available.

9. The proposed Spectral Clustering algorithm proved excellent segmentation results for the whole gland and the boundary detection errors were calculated with 94% average boundary detection accuracy. The algorithm was first tested for boundary detection with excellent ability to identify the boundaries of the gland without any need to pre-specify either an initial contour or a seed point. The algorithm was then applied for segmenting the interior of the gland and the segmented regions did match the radiologist's identified regions. The Spectral Clustering was proposed to tackle the problem from a different point of view that supports the results obtained using the previous methods.
10. A framework for cancer recognition using TRUS images was built with a library full of informative features and intelligent feature selection methods as well as internal gland segmentation methods.

9.3 Future Work

This thesis established a framework for implementing an automated system for identifying suspicious regions, constructing different features, selecting the best features and classifying different regions. However, there is still a need to investigate more avenues in this research such as:

- Extending this work to be applicable for pathology-identified images. It is expected to produce excellent classification results because all the different features contribute to tissue classes.
- Obtaining a bigger data set would be appreciated for accurately judging the constructed features.
- Combining the image features with other clinical data such as the age, family history, PSA level, etc,… is expected to yield fruitful classification results. This is because the information is gathered from different sensors, where different feature fusion methods can be helpful.
- Moreover, extending this work to be integrated in the Ultrasound machine for online cancer detection should be considered, especially with the spectral clustering due to its fast estimates of both the prostate boundary and the internal segmentation of the gland.
- Applying the proposed algorithms for tissue classification using ultrasound images of other body organs such as liver tissue classification and Intra Vascular Ultrasound images (IVUS).
- Generalizing these algorithms to be applied for other image modalities such as the MRI and the CT.

- Applying image registration with different image modalities will also be investigated as each image modality is expected to complement the information gathered from the other image modality.

Appendix A

LEMs Masks

Mask 1= $\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$

Mask 2 = $\begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ -4 & -8 & 0 & 8 & 4 \\ -6 & -12 & 0 & 12 & 6 \\ -4 & -8 & 0 & 8 & 4 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix}$

Mask 3 = $\begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ -4 & 0 & 8 & 0 & -4 \\ -6 & 0 & 12 & 0 & -6 \\ -4 & 0 & 8 & 0 & -4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix}$

Mask 4 = $\begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ 4 & -16 & 24 & -16 & 4 \\ 6 & -24 & 36 & -24 & 6 \\ 4 & -16 & 24 & -16 & 4 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$

Mask 5 = $\begin{bmatrix} -1 & 2 & 0 & -2 & -1 \\ -4 & 8 & 0 & -8 & -4 \\ -6 & 12 & 0 & -12 & -6 \\ -4 & 8 & 0 & -8 & -4 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}$

Mask 6 = $\begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$

Mask 7 = $\begin{bmatrix} 1 & 2 & 0 & -2 & -1 \\ 2 & 4 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -4 & 0 & 4 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix}$

Mask 8 = $\begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 2 & 0 & -4 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix}$

Mask 9 = $\begin{bmatrix} -1 & 4 & -6 & 4 & -1 \\ -2 & 8 & -12 & 8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -8 & 12 & -8 & 2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$

Mask 10 = $\begin{bmatrix} 1 & -2 & 0 & 2 & 1 \\ 2 & -4 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 4 & 0 & -4 & -2 \\ -1 & 2 & 0 & -2 & -11 \end{bmatrix}$

Mask 11 =
$$\begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -4 & -6 & -4 & -1 \end{bmatrix}$$

Mask 12 =
$$\begin{bmatrix} 1 & 2 & 0 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -4 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & -2 & -1 \end{bmatrix}$$

Mask 13 =
$$\begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 1 \end{bmatrix}$$

Mask 14 =
$$\begin{bmatrix} -1 & 4 & -6 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -8 & 12 & -8 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -6 & 4 & -1 \end{bmatrix}$$

Mask 15 =
$$\begin{bmatrix} 1 & -2 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 4 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

Mask 16 =
$$\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ -4 & -16 & -24 & -16 & -4 \\ 6 & 24 & 36 & 24 & 6 \\ -4 & -16 & -24 & -16 & -4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

Mask 17 =
$$\begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ 4 & 8 & 0 & -8 & -4 \\ -6 & -12 & 0 & 12 & 6 \\ 4 & 8 & 0 & -8 & -4 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

Mask 18 =
$$\begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ 4 & 0 & -8 & 0 & 4 \\ -6 & 0 & 12 & 0 & -6 \\ 4 & 0 & -8 & 0 & 4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix}$$

Mask 19 =
$$\begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ -4 & 16 & -24 & 16 & -4 \\ 6 & -24 & 36 & -24 & 6 \\ -4 & 16 & -24 & 16 & -4 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

Mask 20 =
$$\begin{bmatrix} -1 & 2 & 0 & -2 & -1 \\ 4 & -8 & 0 & 8 & 4 \\ -6 & 12 & 0 & -12 & -6 \\ 4 & -8 & 0 & 8 & 4 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}$$

Mask 21 =
$$\begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ 2 & 8 & 12 & 8 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -8 & -12 & -8 & -2 \\ -1 & -4 & -6 & -4 & -1 \end{bmatrix}$$

Mask 22=
$$\begin{bmatrix} 1 & 2 & 0 & -2 & -1 \\ -2 & -4 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 4 & 0 & -4 & -2 \\ 1 & 2 & 0 & -2 & -1 \end{bmatrix}$$

Mask 24 =
$$\begin{bmatrix} -1 & 4 & -6 & 4 & -1 \\ 2 & -8 & 12 & -8 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 8 & -12 & 8 & -2 \\ -1 & 4 & -6 & 4 & -1 \end{bmatrix}$$

Mask 23 =
$$\begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ -2 & 0 & 4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -4 & 0 & 2 \\ 1 & 0 & -2 & 0 & 1 \end{bmatrix}$$

Mask 25 =
$$\begin{bmatrix} 1 & -2 & 0 & 2 & 1 \\ -2 & 4 & 0 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -4 & 0 & 4 & 2 \\ 1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

Bibliography

1. McNeal J.E., “The zonal anatomy of the prostate.” Prostate. 1981; 2(1):35-49.
2. McNeal JE.: “Normal and pathologic anatomy of prostate.” Urology. 1981; 17(3):11-6.
3. American Cancer Society. Cancer facts and figures. 2004.
4. Chang JJ, Shinohara K, Bhargava V, Presti JCJr, “Prospective evaluation of lateral biopsies of the peripheral zone for prostate cancer detection.” Journal of Urology Volume 160: Pages 2111–2114, 1998
5. Barbara Paul, Rajiv Dhir, Douglas Landsittel, Moira R. Hitchens, and Robert H. Getzenberg “Detection of Prostate Cancer with a Blood-Based Assay for Early Prostate Cancer Antigen”. Cancer Research 2005; 65: (10). May 15, 2005
6. Catalona,W.J, Smith, D.S., Ratliff, T.L “ measurement of prostate specific antigen in serum as a screening test for prostate cancer”. N Engl J Med 1991; 324 1156-61
7. Thompson I.M., Pauler, D.K., GoodmanP.J. “Prelevance of prostate canceramong men with a prostate specific antigen level< or = 4.0 ng/ml”. N Engl J Med 2004; 350: 2239-46
8. Standaert, B.; Alwan, A.; Nelen, V.; Denis, L., “Prostate volume and cancer in screening programs”. The Prostate Volume: 33, Issue: 3, 1 November 1997, pp. 188 – 194
9. Hoffelt, S. Christopher; Marshall, Lynn M.; Garzotto, Mark; Hung, Arthur; Holland, John; Beer, Tomasz M. “A comparison of CT scan to transrectal ultrasound-measured prostate volume in untreated prostate cancer”. International Journal of Radiation Oncology*Biology*Physics Volume: 57, Issue: 1, September 1, 2003, pp. 29-32
10. Vapnek JM, Hricak H, Shinohara K, Popovich M, Carroll P. “Staging accuracy of magnetic resonance imaging versus transrectal ultrasound in stages A and B prostatic cancer.”, Urol Int. 1994;53(4):191-5.
11. Siegal JA, Yu E, Brawer MK. “Topography of neovascularity in human prostate carcinoma.” Cancer 1995; 75: 2545 51
12. Smith, Joseph A. “Transrectal ultrasonography for the detection and staging of carcinoma of the prostate”. Journal of Clinical Ultrasound Volume: 24, Issue: 8, October 1996, pp. 455 – 461
13. Presti, JR. JC., O’Dowd GJ, Miller M.C., Mattu R, Veltri RW, “Extended Peripheral Zone Biopsy Schemes Increase Cancer Detection Rates and Minimize Variance in Prostate Specific Antigen and Age Related Cancer Rates: Results of a Community Multi-Practice Study.” The Journal of Urology, 2003 , Vol. 169, Issue 1, Pages 125-129
14. Presti JR JC, Chang JJ, Bhargava V, Shinohara K, “The Optimal Systematic Prostate Biopsy Scheme Should Include 8 Rather Than 6 Biopsies: Results of a Prospective Clinical Trial”. The Journal of Urology January 2000 (Vol. 163, Issue 1, Pages 163-167)
15. Loch T, Leuschner I, Genberg C, et al., “Artificial neural network analysis (ANNA) of prostatic transrectal ultrasound.” Prostate 1999; 39:198–204.
16. Rajesh K Naz, “Prostate; basic and clinical aspects”, CRC Press, 1997,
17. J C Nickel, “Text book of Prostatitis”, Taylor francis (UK), 1999
18. Lenhard Jr., “Clinical Oncology”, Blackwell Publishing, 2001, pp.429
19. Sakti Das, David Crowford, “Cancer of the Prostate”, Marcel Dekker, 1993,pp131

20. Amir V Kaisary, Gerald P Murphy, Louis Denis , "Textbook of Prostate Cancer: Pathology, Diagnosis, and Treatment", Taylor francis (UK), 1999, pp145

21. "The Oncology Channel" http://www.oncologychannel.com/prostatecancer/stagingsystems.shtml

22. Yue J. Wang, Jianhua Xuan, Isabell A. Sesterhenn, Wendelin S. Hayes, David S. Ebert, John H. Lynch, and Seong K. Mun, "Statistical modeling and visualization of localized prostate cancer" SPIE Proceedings, International Society of Optical Engineering, 1997, 3031, 73 (1997)

23. The Dianon Inc. http://www.dianon.com/default.htm

24. Vernon E. Welcon, Hary Neuwirth , Patrick Bennet "Marin Urology" http://www.marinurology.com/index.html

25. John R. Thornbury, David K. Ornstein, Peter L. Choyke, Curtis P. Langlotz and Jeffrey C. Weinreb "Prostate Cancer What Is the Future Role for Imaging?" American Journal of Radiology 2001; 176:17-22

26. Catalona WJ, Hudson MA, Scardino PT, et al. "Selection of optimal prostate specific antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves." Journal of Urology 1994;152:2037 -2042

27. Catalona WJ, Smith DS, Ornstein DK. "Prostate cancer detection in men with serum PSA concentrations of 2.6 to 4.0 ng/mL and benign prostate examination: enhancement of specificity with free PSA measurements." The Journal of American Medical Association 1997;277:1452 - 1455.

28. Arcangeli CG, Ornstein DK, Keetch DW, Andriole GL. "Prostate-specific antigen as a screening test for prostate cancer: the United States experience." North American Clinical Urology 1997;24:299 -306.

29. Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL, "Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men." Journal of Urology 1994,151(5):1308-9.

30. Catalona WJ, Partin AW, Slawin KM, et al. "Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial." JAMA 1998;279:1542 -1547

31. Epstein JI, Walsh PC, Akingba G, Carter HB. The significance of prior benign needle biopsies in men subsequently diagnosed with prostate cancer. J Urol 1999;162:1649 -1652

32. Tillmann Loch , Ursula Eppelmann, Jan Lehmann, Bernd Wullich, Annemie Loch1 and Michael Stöckle, "Transrectal ultrasound guided biopsy of the prostate: random sextant versus biopsies of sono-morphologically suspicious lesions", World Journal of Urology, Volume 22, Number 5, 2004, Pages 1433-8726

33. Levine MA, Ittman M, Melamed J, Lepor H. Two consecutive sets of transrectal ultrasound guided sextant biopsies of the prostate for the detection of prostate cancer. J Urol 1998;159:471 - 475

34. Eskew LA, Bare RL, McCullough DL. Systematic 5 region prostate biopsy is superior to sextant method for diagnosing carcinoma of the prostate. J Urol 1997;157:199 -202

35. Gleason DF, Mellinger GT. "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging." Journal of Urology 1974; 111 (1): 58–64

36. Kotre, J “Image processing in the fight against breast cancer”. Engineering Science and Education Journal , Volume: 2 Issue: 1 , Feb. 1993 Page(s): 41 –46

37. Hyoung Seop Kim; Ishikawa, S.; Ohtsuka, Y.; Shimizu, H.; Shinomiya, T.; Viergever, M.A.; “Automatic scoliosis detection based on local centroids evaluation on moire topographic images of human backs Medical Imaging”, IEEE Transactions on , Volume: 20 Issue: 12 , Dec. 2001 Page(s): 1314 –1320

38. Giansanti, R.; Fumelli, P.; Passerini, G.; Zingaretti, P.; “Imaging system for retinal change evaluation Image Processing and Its Applications”, 1997., Sixth International Conference on , Volume: 2 , 14-17 July 1997 Page(s): 530 -534 vol.2

39. Chiang, H.K.; Chui-Mei Tiu; Guo-Shian Hung; Shiao-Chi Wu; Chang, T.Y.; Yi-Hong Chou; “Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis” Ultrasonics Symposium, 2001 IEEE , Volume: 2 , 7-10 Oct. 2001 Page(s): 1303 -1306 vol.2

40. Chung-Ming Wu; Yung-Chang Chen; Kai-Sheng Hsieh; “Texture features for classification of ultrasonic liver images”. Medical Imaging, IEEE Transactions on, Volume: 11 Issue: 2, June 1992 Page(s): 141 –152

41. Nash, J.M.; Carter, J.N.; Nixon, M.S.; “The velocity Hough transform: a new technique for dynamic feature extraction”. Image Processing, 1997. Proceedings., International Conference on , 26-29 Oct. 1997 Page(s): 386 -389 vol.2

42. Carlier, S.G.; Marijnissen, J.P.A.; Coen, V.L.M.A.; van der Giessen, W.J.; Sabate, M.; Ligthart, J.; den Boer, A.; Cespedes, I.E.; Wenguang Li; van der Steen, A.F.W.; Levendag, P.C.; Serruys, P.W.; “Guidance of intracoronary radiation therapy based on dose-volume histograms derived from quantitative intravascular ultrasound Medical Imaging”, IEEE Transactions on , Volume: 17 Issue: 5 , Oct. 1998 Page(s): 772 –778

43. Bloch, P.; Whittington, R.; Hutchinson, D.M.; Bjarngard, B.; “Application of computed radiography for portal imaging in radiation oncology” Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE, Volume: 4 , 23-28 July 2000 Page(s): 2879 -2885 vol.4

44. R. Gonzalez and R. Woods, Digital Image Processing. Addison-Wesley, 1992.

45. D. Pathak, Peter D. Grimm, Vikram Chalana, and Yongmin Kim “Pubic Arch Detection in Transrectal Ultrasound Guided Prostate Cancer Therapy” IEEE transactions on medical imaging, vol. 17, no. 5, October 1998 pp762-771

46. Coakley, Fergus V.; Qayyum, Aliya; Kurhanewicz, J., “Magnetic Resonance Imaging and Spectroscopic Imaging of Prostate Cancer,” The Journal of Urology: Volume 170(6, Part 2 of 2) December 2003 pp S69-S76

47. Perrotti M, Han KR, Epstein RE, et al. Prospective evaluation of endorectal magnetic resonance imaging to detect tumor foci in men with prior negative prostatic biopsy: a pilot study. J Urol 1999;162:1314 -1317[Medline]

48. Costello LC, Franklin RB, Narayan P. “Citrate in the diagnosis of prostate cancer.” Prostate 1999;38:237 -245.

49. Scheidler J, Hricak H, Vigneron DB, et al. Prostate cancer: localization with three-dimensional proton MR spectroscopic imaging: clinicopathologic study. Radiology 1999;213:473 -480

50. Jager GJ, Ruijter ETG, vanderKaa CA, et al. Dynamic turboFLASH subtraction technique for contrast-enhanced MR imaging of the prostate: correlation with histopathologic results. Radiology 1997;203:645 -652

51. Dresner MA, Rose GH, Rossman PJ, et al. Magnetic resonance elastography of the prostate. (abstr) Radiology 1998;209(P):181

52. Bates T.S., Gillatt D.A.; Cavanagh P.M.; Speakman M. "A comparison of endorectal magnetic resonance imaging and transrectal ultrasonography in the local staging of prostate cancer with histopathological correlation" British Journal of Urology, Volume 79, Number 6, June 1997, pp. 927-932(6)

53. M.R.W. Engelbrecht, J.O. Barentsz, G.J. Jager, M. Van Der Graaf, A. Heerschap, J.P.M. Sedelaar, R.G. Aarnink and J.J.M.C.H. De La Rosette, "Prostate cancer staging using imaging", BJU International (2000), 86 Suppl. 1, 123±134

54. Hinkle GH, Burgers JK, Neal CE, et al. Multicenter radioimmunoscintigraphic evaluation of patients with prostate carcinoma using indium-111 capromab pendetide. Cancer 1998;83:739 -747

55. Chang SS, Bander NH, Heston WD. Monoclonal antibodies: will they become an integral part of the evaluation and treatment of prostate cancer-focus on prostate-specific membrane antigen? Curr Opin Urol 1999;9:391 -395

56. Seltzer MA, Barbaric Z, Belldegrun A, et al. Comparison of helical computerized tomography, positron emission tomography and monoclonal antibody scans for evaluation of lymph node metastases in patients with prostate specific antigen relapse after treatment for localized prostate cancer. J Urol 1999;162:1322 -1328

57. Lattanzi J.1; McNeeley S.; Pinover W.; Horwitz E.; Das I.; Schultheiss T.E.; Hanks G.E." A comparison of daily CT localization to a daily ultrasound-based system in prostate cancer - further evidence for a significant dose effect ": International Journal of Radiation Oncology*Biology*Physics, Volume 43, Number 4, 1 March 1999, pp. 719-725(7)

58. VS Khoo, AR Padhani, SF Tanner, DJ Finnigan, MO Leach and DP Dearnaley, "Comparison of MRI with CT for the radiotherapy planning of prostate cancer: a feasibility study", The British Journal of Radiology, Vol 72, Issue 858 590-597,

59. Effert, Peter J.; Bares, Roland; Handt, Stefan; Wolff, Johannes M.; Bull, Udalrich; Jakse, Gerhard," Metabolic Imaging of Untreated Prostate Cancer by Positron Emission Tomography with sup 18 Fluorine-Labeled Deoxyglucose", The Journal of Urology: Volume 155(3) March 1996 pp 994-998

60. Aarnink RG, Beerlage HP, De la Rosette JJMCH, et al. Transrectal ultrasound of the prostate: innovations and future applications. J Urol 1998;159:1568 -1579

61. Patel U, Rickards D. The diagnostic value of colour Doppler ¯ow in the peripheral zone of the prostate with histological correlation. Br J Urol 1994; 74: 590±5 36 Cornud F, Belin

62. Lorenz A, Sommerfeld HJ, Garcia-Schu¨rmann M, et al. "Diagnosis of prostate carcinoma using multicompression strain imaging Data acquisition and first in vivo results." IEEE Ultrason Sympos 1998: 1761–1764.

63. Lorenz A, Sommerfeld HJ, Garcia-Schu¨rmann M, et al. "A new system for the acquisition of ultrasonic multicompression strain images of the human prostate in vivo." IEEE Transactions Ultrasonics Ferroelec Freq Control 1999;46:1147–1153.

64. Pesavento A, Perrey C, Krueger M, Ermert H. "A time-efficient, and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation". IEEE Trans Ultrasonics Ferroelec Freq Control 1999b; 46:1057–1066.

65. Pesavento A, Lorenz A. "Real time strain imaging and in-vivo applications in prostate cancer". IEEE Transactions on Ultrasonics Symposium 2001; 2:1647–1652.

66. Lizzi, F. L.; Feleppa, E. J.; Kaisar A., S.; Deng, C. X.; "Ultrasonic spectrum analysis for tissue evaluation"; Pattern Recognition Letters Volume: 24, Issue: 4-5, February, 2003, pp. 637-658

67. Feleppa, E.J.; Kalisz, A.; Sokil-Melgar, J.B.; Lizzi, F.L.; Tian L.; Rosado, A.L.; Shao, M.C.; Fair, W.R.; Yu W.; Cookson, M.S.; Reuter, V.E.; Heston, W.D.W.; "Typing of prostate tissue by ultrasonic spectrum analysis" IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, , Volume: 43, Issue: 4, July 1996, Pages: 609 – 619

68. Lizzi, F.L; Feleppa, E.J.; Astor, M.; and Kalisz, A.; "Statistics of Ultrasonic spectral parameters for prostate and liver examinations" ; IEEE Transactions on Ultrasonics, Ferroelectrics and frequency control, Volume: 44 , 7-July 1997 Pages: 935-942

69. Bassat O., Sun Z., Mestas J.L., Gimenez G., "Texture Analysis Of Ultrasound Images of the Prostate By Means of Co-occurrence Matrices", Ultrasonic Imaging 15, 1993 ,Pages: 218-237

70. Scheipers, Ulrich; Ermert, Helmut; Sommerfeld, Hans-Joerg; Garcia-Schürmann, Miguel; "Ultrasonic multifeature tissue characterization for prostate diagnostics"; Ultrasound in Medicine and Biology Volume: 29, Issue: 8, August, 2003, pp. 1137-1149

71. S S Mohamed, M M A Salama, M Kamel, E F El-Saadany, K Rizkalla and J Chin, "Prostate cancer multi-feature analysis using trans-rectal ultrasound images" Phys. Med. Biol. 50 No 15 (7 August 2005) N175-N185

72. Dinggang Shen; Yiqiang Zhan; Davatzikos, C.;Medical Imaging, "Segmentation of prostate boundaries from ultrasound images using statistical shape model" IEEE Transactions on , Volume: 22 , Issue: 4 , April 2003 Pages:539 – 551

73. Lixin Gong; Pathak, S.D.; Haynor, D.R.; Cho, P.S.; Yongmin Kim; "Parametric shape modeling using deformable superellipses for prostate segmentation" Medical Imaging, IEEE Transactions on , Volume: 23, Issue: 3, March 2004 Pages:340 – 349

74. David A. Clausi, M.Ed Jernigan, Designing Gabor filters for optimal texture separability, Pattern Recognition 33 (2000) 1835-1849.

75. Anil K. Jain, Farshid Farrokhnia, Unsupervised texture segmentation using Gabor filters, Pattern Recognition, vol. 24,no.12, pp 1167-1186,1991.

76. A.C. Bovik, M. Clark, W.S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Trans. Pattern Anal. Machine Intell. 12 (1) (1990) 55-73.

77. Weldon, Thomas P.; Higgins, William E.; Dunn, Dennis F., " Efficient gabor filter design for texture segmentation', Pattern Recognition Volume: 29, Issue: 12, December, 1996, pp. 2005-2015

78. Mehrotra, R.; Namuduri, K.R.; Ranganathan, "Gabor filter-based edge detection" pattern recognition vol.25 issue 12 December 1992 N. pp. 1479-1494

79. Daugman, J.G.; "High confidence visual recognition of persons by a test of statistical independence" Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 15, Issue 11, Nov. 1993 Page(s):1148 - 1161

80. Thomas P. Weldon, William E. Higgins, and Dennis F. Dunn "Gabor filter design for multiple texture segmentation" Opt. Eng. 35, 2852 (1996)

81. Jain, Anil K.; Ratha, Nalini K.; Lakshmanan, Sridhar, "Object detection using Gabor filters", Pattern Recognition Volume: 30, Issue: 2, February, 1997, pp. 295-309

82. J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Opt. Soc. Am. A 2, 1160- (1985)

83. P. Brodatz. " Textures: A Photographic Album for Artists and Designers"

84. Shen, Dinggang; Lao, Zhiqiang; Zeng, Jianchao; Zhang, Wei; Sesterhenn, Isabel A.; Sun, Leon; et. al. "Optimized prostate biopsy via a statistical atlas of cancer spatial distribution", Medical Image Analysis Volume: 8, Issue: 2, June, 2004, pp. 139-150

85. Schwartz L. "Imaging and prostate cancer". J. Radiation Oncology 36(1-supplement):118:1996.

86. Clements R. "Has ultrasonography a role for screening for prostatic cancer?" Europian Radiolology 7:217-223:1997.

87. M. Nixon and A. Aguado," Feature extraction and image processing". Newnes, 2002

88. W. Pratt, "Digital Image Processing". New York: John Wiley, 1991.

89. Konstantinos I Papastathopoulos and Jost B Jonas, "Follow up of focal narrowing of retinal arterioles in glaucoma," British Journal of Ophthalmology., Mar 1999; 83: 285 - 289.

90. Chen C.H., Pau L.F., Wang P.S.P., The Handbook of Pattern Recognition and Computer Vision (2nd Edition), World Scientific Publishing Co., 1998.

91. B. Juleze, "Experiments in the visual perception of texture," Scientif. Amer., pt. 4, vol. 232, pp. 34–43, 1975.

92. Haralick, R.M., K. Shanmugam and I. Dinstein. 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics. SMC-3(6):610-620

93. Haralick, R.M., "Statistical and structural approaches to texture" Proceedings of the IEEE, Volume 67, Issue 5, May 1979 Page(s):786 – 804

94. J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," IEEE Trans. Systems, Man, Cybern., vol. SMC-6, pp. 269-285, 1976.

95. Peter A. Freeborough* and Nick C. Fox, "MR Image Texture Analysis Applied to the Diagnosis and Tracking of Alzheimer's Disease" IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 17, NO. 3, JUNE 1998 475

96. JE Koss, FD Newman, TK Johnson, DL Kirch, "Abdominal organ segmentation using texture transforms and a Hopfield neural network", IEEE Transactions on Medical Imaging, Volume. 18, No. 7, July 1999, Pages 640-649

97. Leen-Kiat Soh, Costas Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 37, NO. 2, MARCH 1999

98. DANIELLE J. MARCEAU, PHILIP J. HOWARTH, JEAN-MARIE M. DUBOIS, AND DENIS J. GRATTON "Evaluation of the Grey-Level Co-Occurrence Matrix Method For Land-Cover Classification Using SPOT Imagery" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 28, NO. 4, JULY 1990 513

99. Clausi, D. A.. An analysis of co-occurrence texture statistics as a function of grey-level quantization. Canadian Journal of Remote Sensing, 2002 vol. 28 no. 1 pp. 45-62

100. David A. Clausi and M. Ed Jernigan "A Fast Method to Determine Co-Occurrence Texture Features" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 36, NO. 1, JANUARY 1998

101. K. Bhanu Prakash, A. Ramakrishnan; " Fetal Lung Maturity Analysis Using Ultrasound Image Features', IEEE Transactions on Information Technology in Biomedicine, Vol.6, No.1, march 2002.

102. Jose´e Le´vesquea, Douglas J. King, "Spatial Analysis of radiometric fractions from high resolution multispectral imagery for modeling individual tree crown and forest canopy structure and health" Remote Sensing of Environment, 2003, Remote Sensing of Environment 84 (2003) 589–602

103. Augusteijn, M.F.; Clemens, L.E.; Shaw, K.A.;" Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier", IEEE Transactions on Geoscience and Remote Sensing ,Volume 33, Issue 3, May 1995 Page(s):616 - 626

104. M. Amadasun and R. King, "Textural features corresponding to textural properties", IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 5, pp. 1264-1274, September 1989.

105. M. M. Galloway, "Texture analysis using gray-level run lengths", Computer Graphics and Image Processing, vol. 4, pp. 172-179, June 1975.

106. Hart P., "The condensed nearest neighbour rule" IEEE Transactions on Information Theory, Volume: 14 , Issue: 3 , May 1968 Pages:515 – 516

107. Kraaijveld, "An Experimental Comparison of Non-Parametric Classifiers for Time-Constrained Classification Tasks",ICPR98, Vol I,pp: 428-435

108. Quinlan, J.R. "Decision trees and decision-making" IEEE Transactions on Systems, Man and Cybernetics, Volume: 20, Issue: 2, Mar/Apr 1990, Page(s): 339-346

109. Duda R. O., P.E.Hart, D.G.Stork; "Pattern Classification", John Wiley&Sons Inc. 2001.

110. Boser, B.E., Guyon, I., & Vapnik, V.N., "A training algorithm for optimal margin classifiers". Proceedings of the Fifth Annual Workshop of Computational Learning Theory, 5, 144-152, Pittsburgh, ACM.

111. Cortes, C.; Vapnik, V. "Support-Vector Networks", Machine Learning Volume: 20, Issue: 3, September 1995, pp. 273-297

112. Kadah, Y.M.; Farag, A.A.; Zurada, J.M.; Badawi, A.M.; Youssef, A.-B.M.;Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images, IEEE Transactions on Medical Imaging Volume 15, Issue 4, Aug. 1996 Page(s):466 – 478

113. Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Statistics

114. S.F. Arnold. Mathematical Statistics. 1990, Prentice-Hall.

115. R.B. D'Agostino and M.A. Stephens, Eds. (1986). Goodness-of-Fit Techniques, Marcel Dekker.

116. The Minitab software website "http://www.minitab.com/education/"

117. D. Hand, H. Mannila, and P. Smyth. "Principles of Data Mining". A Bradford Book The MIT press, 2001.

118. H. Liu and H. Motoda, "Feature Extraction, Construction and Selection: A Data Mining Perspective". Boston: Kluwer Academic Publishers, 1998. 2nd Printing, 2001

119. P. Langley "Elements of Machine Learning". Morgan Kaufmann, 1996.

120. J.Novovicova P.Somol,P.Pudil and P.Paclik. "Adaptive floating search methods in feature selection". Pattern Recognition Letter 20(11-13):1157 .1163, 1999

121. Guo-Zheng Li1, Jie Yang1, Guo-Ping Liu1 and Li Xue1, "Feature Selection for Multi-class Problems Using Support Vector Machines", PRICAI 2004: Trends in Artificial Intelligence:

Lecture Notes in Computer Science Publisher: Springer Berlin / Heidelberg, Volume 3157 / 2004, pp: 292 - 300

122. Battiti, R, "Using mutual information for selecting features in supervised neural net learning.", IEEE Transactions on Neural Networks, Volume: 5, Issue: 4, Jul 1994 Page(s): 537-550

123. Grall-Maes, E.; Beauseroy, P. "Mutual information-based feature extraction on the time-frequency plane"; IEEE Transactions on Signal Processing Volume 50, Issue 4, April 2002 Page(s):779 - 790

124. N. Kwak and C.-H. Choi, "Input Feature Selection for Classification Problems," IEEE Transactions Neural Networks, vol. 13, no. 1, pp. 143-159, Jan. 2002.

125. Kwak, N.; Chong-Ho Choi; "Input feature selection by mutual information based on Parzen window" IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 24, Issue 12, Dec. 2002 Page(s):1667 – 1671

126. Luis Talavera, "An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering" Lecture Notes in Computer Science Publisher: Springer Berlin / Heidelberg, Volume 3646 / 2005, Page: 440:451

127. Kennedy, J.; Eberhart, R.; "Particle swarm optimization" Neural Networks, 1995. Proceedings., IEEE International Conference on Volume 4, 27 Nov.-1 Dec. 1995 Page(s):1942 - 1948 vol.4

128. Eberhart, R.C.; Shi, Y.; "Guest Editorial Special Issue on Particle Swarm Optimization" IEEE Transactions on Evolutionary Computation, Volume 8, Issue 3, June 2004 Page(s):201 – 203

129. Dorigo, M. Maniezzo, V. and Colorni, A. "The ant system: optimization by a colony of cooperating agents", IEEE Transactions on Systems, Man, and Cybernetics-Part B 26(1) (1996), pp. 29–41.

130. Coggins J.M. and Jain, A.K., "A Spatial Filtering Approach to Texture Analysis", Pattern Recognition Letters, Volume: 3, No. 3, Pages: 195-203, 1985.

131. Randen, T.; Husoy, J.H., "Filtering for texture classification: a comparative study.", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 21, Issue 4, April 1999 Pages:291 – 310

132. Proakis, J.G., and Manolakis D.G., Digital Signal Processing: Principles, Algorithms, and Applications. Englewood Cliffs, NJ: Prentice Hall, 1996.

133. Hayes, M.H., Statistical Digital Signal Processing and Modeling. New York: John Wiley & Sons, 1996.

134. Gupta, I.J., Beals, M.J ,and Moghaddar, A., "Data extrapolation for high resolution radar imaging.", IEEE Transactions on Antennas and Propagation , Volume 42, Issue 11, Pages: 1540 – 1545, Nov. 1994

135. Vizinho A., and Wyatt L.R., "Evaluation of the use of the modified-covariance method in HF radar ocean measurement.", IEEE Journal of Oceanic Engineering, Volume 26, Issue 4, Pages. 832 - 840, 2001.

136. the NASA website http://southport.jpl.nasa.gov/

137. Welch, P.D. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms." IEEE Trans. Audio Electroacoustics Vol. AU-15 (June 1967). Pgs. 70-73.

138. Matlab, Signal Processing Toolbox documentation

139. Thomson, D.J., "Spectrum estimation and harmonic analysis," Proceedings of the IEEE, Vol. 70 (1982), pp. 1055-1096

140. McCoy, E.J., Walden, A.T., Percival, D.B. " Multitaper spectral estimation of power law processes Signal Processing.", IEEE Transactions on Acoustics, Speech, and Signal Processing, Volume 46, Issue 3, March 1998 Pages: 655 – 668

141. Burrows, M.L., "Two-dimensional ESPRIT with tracking for radar imaging and feature extraction", IEEE Transactions on Antennas and Propagation, Volume 52, 2004, Pages:524 – 532

142. Ottersten, B., Viberg, M., Kailath, T., "Performance analysis of the total least squares ESPRIT algorithm", IEEE Transactions on Signal Processing, Volume 39, 1991 Pages:1122 – 1135

143. Roy, R., and kailath, T., "Esprit Estimation of signal parameters via rotational invariance techniques", IEEE Transactions on Acoustics, Speech and Signal processing, Volume 37, 1998, Pages: 984-995

144. Roy, R., Paulraj, A., and Kailath, T., "ESPRIT- A Substance rotation approach to estimation of parameters of cissoids in noise", IEEE Transactions on Acoustics, Speech and Signal Processing. 1986, Volume 34,

145. Cross S., "Fractals in Pathology", The Journal of Pathology, Volume 182, Issue 1 (p 1-8)

146. Pentland P., "Fractal-based description of natural scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 661-674, November 1984.

147. Sarkar. N and Chaudhuri B. "An efficient differential box counting approach to compute fractal dimension of an image", IEEE Trans. Systems man and Cybernetics, vol. 24, No. 1, January 1994.

148. Wen-Li Lee; Yung-Chang Chen; Kao-Sheng Hsieh; "Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform", IEEE Transactions on Medical Imaging, vol. 22 , March 2003 pp. 382-391.

149. Gache, N.; Flandrin, P.; Garreau, D.; "Fractal dimension estimators for fractional Brownian motions", International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp.3557 – 3560, April 1991.

150. Penn, A.I.; Loew, M.H.; "Estimating fractal dimension with fractal interpolation function models" IEEE Transactions on Medical Imaging, Volume 16, Issue 6, Dec. 1997 Page(s):930 - 937

151. Laws K. "Rapid texture identification". In SPIE Image Processing for Missile Guidance, Vol. 238, pp. 376-380, 1980.

152. Roweis S. and Saul L "Nonlinear dimensionality reduction by locally linear embedding" Science, vol. 290 no. 5500, pp.2323-2326, Dec. 2000.

153. Abolmaesumi, P., Sirouspour, M.: "An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images." IEEE Transactions on Medical Imaging 23 (2004) 772–784

154. Shen, D., Zhan, Y., Davatzikos, C.: "Segmentation of the prostate boundaries from ultrasound images using statistical shape model". IEEE Transactions on Medical Imaging 22 (2003) 539–551

155. Ng, A.Y., Jordan, M., Weiss, Y.: "On spectral clustering: Analysis and an algorithm": Advances in Neural Information Processing Systems 14, MIT Press (2002) 849– 856

156. Shi, J., Malik, J.: "Normalized cuts and image segmentation". IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 888–905

157. Timothee Cour, Florence Benezit, Jianbo Shi. “Spectral segmentation with multiscale graph decomposition” IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Volume 2, June 2005, Page 1124 - 1131.

158. Perona and W. T. Freeman. “A factorization approach to grouping”. In H. Burkardt and B. Neumann, editors, Proc ECCV, pages 655-670, 1998.

159. J. Shi and J. Malik. “Normalized cuts and image segmentation”. IEEE Conf. on Computer Vision and Pattern Recognition, pages 731-737, 1997.

160. N. Archip, R. Rohling, P. Cooperberg, H. Tahmasebpour, and S. K. Warfield, “Spectral Clustering Algorithms for Ultrasound Image Segmentation”, MICCAI 2005, LNCS 3750, pp. 862–869.

List of Publications

Accepted Journal Papers:

1. S.S. Mohamed, M.M.A. Salama, M. Kamel, E.F El-Sadaany, K.Rizkalla ,J.Chin “Prostate cancer multi-feature analysis using TRUS images” Physics in Medicine and Biology : 50 NO 15 (7 August 2005) N175-N185
2. S. S. Mohamed , A. M. Youssef , E. F. EL-Sadaany , and M. M. A. Salama , "LLE Based TRUS Image Features Dimensionality Reduction for Prostate Cancer Diagnosis", International Journal on Graphics, Vision and Image Processing (in Press)

Submitted Journal Papers:

1. S.S. Mohamed, M. M.Salama and K. Rizkalla “Spectral Clustering for TRUS Image segmentation” Submitted to Journal of Electronic Imaging
2. S.S. Mohamed and M. M.Salama “Prostate cancer Spectral Multi-Feature Analysis using TRUS Images” Submitted to Physics in Medicine and Biology

Refreed Book Chapters:

1. S.S. Mohamed, A.M.Youssef, E.F El-Saadany, M.M.A. Salama "Artificial Life Feature Selection Techniques for Prostrate Cancer Diagnosis Using TRUS Images" Lecture Notes in Computer Science, ICIAR 2005, LNCS 3656, pp. 903 – 913.
2. S.S. Mohamed, M.M.A. Salama, M. Kamel, K. Rizkalla “Region of interest based prostate tissue characterization using least square support vector machine LS-VSM”, Lecture Notes in Computer Science, ICIAR 2004, Vol. 3212, pp. 51-58
3. S.S. Mohamed, A.M.Youssef, E.F El-Saadany, M.M.A. Salama “Prostate Tissue Characterization using TRUS Image Spectral Features” Lecture Notes in Computer Science, ICIAR 2006, In Press

Refreed Conference Papers:

1. S.S. Mohamed, M. M. A. Salama, “Computer-aided diagnosis for prostate cancer using support vector machine” Publication: Proc. SPIE Vol. 5744, p. 899-907, Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display; Robert L. Galloway, Jr.,Kevin R. Cleary; Eds.
2. S. S. Mohamed, A. M. Youssef, E. F. EL-Sadaany, and M. M. A. Salama, “AIS TLS-ESPRIT Feature Selection for Prostate Tissue Characterization". Publication: Proc. SPIE Vol. 6144, 614451, pp 614451-1 through 614451-11: Medical Imaging 2006: Image Processing; Robert L. Galloway, Jr.,Kevin R. Cleary; Eds.

3. S.S. Mohamed, M.M.A. Salama "Higher order Texture Feature Assesment for Prostate Cancer Diagnosis", 8th International Conference on Pattern Recognition and Information Processing (PRIP'05), pp 152-155, May 18-20, 2005, Minsk, Belarus
4. S.S. Mohamed, M.M.A. Salama, and G. Freeman "Texture Feature Analysis for Prostate Cancer Diagnosis". The 2nd Cairo International Biomedical Engineering Conference, CIBEC 2004. Dec. 27 to 2004 - Dec. 29, 2004

Conference Papers:

1. S.S. Mohamed, E.F. El-Sadaany, J. Shen, T.K. Abdelgalil, , M.M.A. Salama, D.B. Downey, K. Rizkalla, A.Finster, Dec. 2003, "Region of interest identification in prostate TRUS images based on Gabor filter", The 46th Midwest Symposium on Circuits and Analysis, Cairo, Egypt, 27-30 December 2003 Vol. 1 pp. 415: 419
2. S.S. Mohamed, T. K. Abdel-galil, M. M. A. Salama, A. Fenster, , K. Rizkalla and D. B. Downey, May 2003 "Prostate Cancer Diagnosis Based on Gabor Filter Texture Segmentation of Ultrasound Image", IEEE Canadian Conference on Electrical and Computer Engineering (CCECE'03), Montreal, Canada, Vol. 3 , pp. 1485 –1488.
3. S. S. Mohamed, A. M. Youssef, E. F. EL-Sadaany, and M. M. A. Salama, "On the use of spectral clustering for prostate tissue charecterizzation" IEEE Canadian Conference on Electrical and Computer Engineering (CCECE'06), Ottawa, Canada,, pp.