IMAGE TEXTURE ANALYSIS OF TRANSVAGINAL ULTRASOUND IN MONITORING OVARIAN CANCER

A thesis submitted to the Cardiff University
in candidature for the degree of
Doctor of Philosophy

By
Bidi Ab. Hamid

Institute of Medical Engineering
and Medical Physics,
Cardiff School of Engineering,
Cardiff University.

December 2011

Abstract

Ovarian cancer has the highest mortality rate of all gynaecologic cancers and is the fifth most common cancer in UK women. It has been dubbed "the silent killer" because of its non-specific symptoms. Amongst various imaging modalities, ultrasound is considered the main modality for ovarian cancer triage. Like other imaging modalities, the main issue is that the interpretation of the images is subjective and observer dependent. In order to overcome this problem, texture analysis was considered for this study. Advances in medical imaging, computer technology and image processing have collectively ramped up the interest of many researchers in texture analysis. While there have been a number of successful uses of texture analysis technique reported, to my knowledge, until recently it has yet to be applied to characterise an ovarian lesion from a B-mode image. The concept of applying texture analysis in the medical field would not replace the conventional method of interpreting images but is simply intended to aid clinicians in making their diagnoses.

Five categories of textural features were considered in this study: grey-level co-occurrence matrix (GLCM), Run Length Matrix (RLM), gradient, auto-regressive (AR) and wavelet. Prior to the image classification, the robustness or how well a specific textural feature can tolerate variation arises from the image acquisition and texture extraction process was first evaluated. This includes random variation caused by the ultrasound system and the operator during image acquisition. Other factors include the influence of region of interest (ROI) size, ROI depth, scanner gain setting, and 'calliper line'. Evaluation of scanning reliability was carried out using a tissue-equivalent phantom as well as evaluations of a clinical environment.

Additionally, the reliability of the ROI delineation procedure for clinical images was also evaluated. An image enhancement technique and semi-automatic segmentation tool were employed in order to improve the ROI delineation procedure. The results of the study indicated that two out of five textural features, GLCM and wavelet, were robust. Hence, these two features were then used for image classification purposes.

To extract textural features from the clinical images, two ROI delineation approaches were introduced: (i) the textural features were extracted from the whole area of the tissue of interest, and (ii) the anechoic area within the normal and malignant tissues was excluded from features extraction. The results revealed that the second approach outperformed the first approach: there is a significant difference in the GLCM and wavelet features between the three groups: normal tissue, cysts, and malignant. Receiver operating characteristic (ROC) curve analysis was carried out to determine the discriminatory ability of textural features, which was found to be satisfactory.

The principal conclusion was that GLCM and wavelet features can potentially be used as computer aided diagnosis (CAD) tools to help clinicians in the diagnosis of ovarian cancer.

Acknowledgements

I offer my deep and sincere thanks to my supervisors: Professor Len Nokes - I will be forever influenced by his incisive ideas, encouragement, and clarity of thought, and Dr Neil Pugh - whose constant guidance and insights were invaluable and always stimulating. I could not wish for better supervisors. Heartfelt thanks, too, to Dr Declan Coleman, for his priceless advice on the technical aspects of texture analysis, and infinite patience in helping me with the scanning and reading of my draft.

I would like to thank Dr Mary Jones, from the School of Mathematics, for her advice concerning statistical matters. I would also like to thank Paul Williams, Kate Wells, and Rhys Morris, for their assistance on technical aspects of ultrasound. I offer my kind regards to the staff at the Doppler Ultrasound Department for directly and indirectly extending their help during the work. I also offer my regards to my friends and colleagues who supported me in all respects as the project progressed.

My very special thanks go to my wife, Sal – you constantly inspired me to better myself and aim higher. Also to my kids: Carmiela, Camelia, Firdaus, and Faiz. To them I dedicate this thesis.

Lastly, the financial support of the Government of Malaysia, who granted my studies, is gratefully acknowledged.

Contents

Appendices

List of Tables

List of Figures

Abbreviation

2D	Two-dimensional
3D	Three-dimensional
ACS	American Cancer Society
AHRQ	Agency for Healthcare Research and Quality
AR	Auto-regressive
AUC	Area under the curve
CMI	Cardiff Malignancy Index
CoV	Coefficient of variation
CT	Computerised tomography
EUS	Endoscopic ultrasound
FIGO	International Federation of Gynecology and Obstetrics
GLCM	Grey level co-occurrence matrix
IOTA	International Ovarian Tumour Analysis group
MRI	Magnetic Resonance Imaging
NVB	Neuro-vascular bundle
PI	Pulsatility Index
Pixel	Picture cell or element
PSV	Peak Systolic Velocity
RCOG	Royal College of Obstetricians and Gynaecologists
RF	Radio-frequency
RI	Resistive Index
RLM	Run length matrix
RMI	Risk Malignancy Index
ROC	Receiver operating characteristic
ROI	Region of interest
SD	Standard deviation
TGC	Time gain compensation
TSV	Transvaginal

Glossary

Amplitude	The maximal distance a particle moves away from its original position.
Area of interest	The area in the image that encompasses the area of the tissue to be characterised.
Computer-aided diagnosis (CAD)	A diagnostic aid that takes into account equally the role of the physician and the benefits of the computer system.
Dynamic range	The range of echoes processed and display by the system, from strongest to weakest.
Frequency	A number of times a vibrating particle goes through its original position within one second.
Gain	The degree of amplification applied to all returning signals.
Lead zirconate titanate (PZT)	The material from which the piezoelectric plate of many transducers is constructed.
Region of interest (ROI)	A region used to extract texture features.
Repeatability	The closeness of the agreement between successive measurements of the object carried out under the same conditions of measurement.
Reproducibility	The closeness of the agreement between measurements of the object carried out under changed conditions of measurement.
ROI depth	Vertical position of the centre of ROI drawn in the image.
ROI size	The area of the drawn region for textural features derivation, measured in pixels.
Sensitivity	Proportion of true positives correctly identified as such.
Specificity	Proportion of true negatives correctly identified as such.
Time gain compensation (TGC)	Use to compensate for the effects of attenuation by progressively increasing the amount of amplification applied to signals with depth, to produce an image of uniform brightness from top to bottom.
Wavelength	The distance between the two adjacent wave peaks.

Chapter 1
Introduction

1.1 Overview

Ovarian cancer has the highest mortality rate of all gynaecologic cancers (Jeong *et al.* 2000; Fishman *et al.*, 2001; Chan and Selman 2006; ACS, 2007; Cancer Research UK, 2010(a)) and has been dubbed "the silent killer" because of its non-specific symptoms (Chan and Selman, 2006). In the United Kingdom, nearly 7,000 new ovarian cancer cases are diagnosed every year (Cancer Research UK, 2009(a)), which is the fifth most common cancer in females and the second most common gynaecological cancer after uterus (Cancer Research UK, 2010(a)). It has been reported that worldwide, around 140,000 women died of ovarian cancer in 2008 (Cancer Research UK, 2010(a)). Its incidence has been steadily increasing over the past ten years, with an overall lifetime risk of 1.8% (Fishman *et al.*, 2001).

Discriminating benign from malignant cases is important, not only to ensure appropriate management for malignant cases (Valentin *et al.*, 2006), but also to avoid unnecessary diagnostic procedures, including surgery, for the non-malignant cases (AHRQ, 2006).

According to Kinkel *et al.* (2000), ultrasound is the primary imaging modality in the evaluation of ovarian masses and is the main triage method prior to treatment. Even though it cannot necessarily prevent the surgery, it helps to narrow down the differential diagnosis and determining the degree of suspicion for malignancy, usually together with the serum CA-125 level (Togashi, 2003). However, the accuracy of

ultrasound imaging based on the morphology assessment is limited due to the significant number of false-positive results (Kupesic and Plavsic, 2006) and raises the issue of how to correctly interpret the images (Gramellini *et al.*, 2008). It has been reported that the suboptimal specificity of grey-scale scoring systems can result in many unnecessary operations when an asymptomatic population is screened for ovarian cancer (Fishmen *et al.*, 2001). This is not surprising because, according to Chan and McCarty (1990), to visually identify a specific tissue is not an easy task and there is always a possibility of human error. Furthermore, according to Giger *et al.* (2008), limitations in the human eye-brain visual system, reader fatigue, distraction, and the presence of overlapping structures that camouflage disease in images may cause detection and interpretation errors. For instance, according to Doi (2007), a radiologist has difficulties in the detection of lung nodules due to the overlap of the normal anatomic structures with nodules. Moreover, since interpreting an ultrasound image with the human eye is always subjective (White *et al.,* 1997), the interpretation of the images is highly dependent on the ability and experience of the observer (Wang *et al.*, 2002; Shung, 2006; Huang *et al.*, 2008; Rocha *et al.*, 2011). This raises the issue of accuracy and reproducibility.

In recent years, objective methods have been proposed to reduce the subjectivity and operator-dependence, but there is no reliable technique available at present. Hence, a new objective method is required to address the above mentioned issues which will contribute to the patients' management.

Texture analysis, a technique for evaluating the position and intensity of the pixels, provides an objective measure to quantify the image properties.

> *'Texture analysis of medical images is a sophisticated computer-assisted technique that allows detection of mathematical patterns in the grey-level distribution of the pixels of digital images, thus providing an objective description of the signal behaviour of anatomic structures or pathological processes'* (Mayerhoefer *et al.*, 2008).

According to Tuceryan and Jain (1998), texture analysis has played a major role in some domain such as remote sensing, while in the medical domain, comprehensive research has been carried out, especially over the past two decades (Doi *et al.*, 1999; Giger *et al.*, 2008).

Results from the image texture analysis technique can be used as an input for computer-aided diagnosis (CAD). According to Dhawan (2003), image analysis techniques have been explored for CAD to improve the sensitivity and specificity of radiological tests. Lack of reproducibility due to subjective interpretation can be reduced by using a computer method (Smyth *et al.*, 1997). According to Giger *et al.* (2008), the potential use of computers has been considered for the analysis of radiographic images since the 1950s.

In texture analysis, the variation in intensity reflects some physical variation in the underlying scene (Davis, 1980). According to Szczypinski *et al.* (2009), image texture of medical images describes internal structure of human tissues or organs. It also describes pathological changes (Mayerhoefer et al., 2008; Xian, 2010). The texture analysis of ultrasound images relies on the principles that, if disease processes affect the structure of the tissue, the tissue should reflect an altered ultrasound signal, which will in turn yield texture feature values different to the normal tissue (Morris, 1988(b)).

There has been a significant advance in medical imaging in the last two decades due to the collective contributions from many areas of medicine, engineering, and basic sciences (Dhawan, 2003). Diagnostic ultrasound image quality in particular has greatly improved. For instance, developments in transducer design have resulted in transducers with greater sensitivity (Averkiou *et al.*, 1997) while broadband transducers lead to improvements in axial resolution (Browne *et al.,* 2004). In addition, the introduction of new techniques such as tissue harmonic imaging (Claudon *et al.,* 2002) and advances in computer technology (Strub, 2005) have also improved the quality of B-mode images.

1.2 Research Hypothesis

Based on current developments in ultrasound imaging, computer technology, image processing, and extensive research works on texture analysis involving medical images, it is hypothesized that the texture analysis technique can be used to characterise and quantify ovarian tissue based on B-mode image texture.

The motivation for this technique is clear: since the output of texture analysis of an image can be expressed numerically, it provides a quantitative means of image description, which in turn, may be able to eliminate the subjectivity. This technique would also solve the reproducibility issue encountered by the conventional image interpretation method. A combination of knowledge, the cognitive capability of humans, and results from a computer is expected to improve the overall diagnosis.

1.3 Aim and Objectives

The aim of this study was to develop an objective method to be used in an ultrasound-based assessment of ovarian tumours, using texture analysis techniques on grey-scale transvaginal (TSV) ultrasound images, with a view to providing a tool that is able to assist the clinician in their diagnosis of ovarian cancer.

In texture analysis, it is important to ensure that the differences in extracted textural features are due to the differences in the acoustic properties of reflecting material and not due to the deterioration of the system or process component. Similarly, a series of scans from the same object must be reproducible. Therefore, the variations in the textural features computed due to image acquisition and features' extraction processes need to be assessed. Hence, in achieving the above aim, the following objectives have been set:

- To evaluate factors that may influence the extracted texture features.
- To evaluate the effects of ultrasound settings/parameters on textural features.
- To assess the ability of the ultrasound system and the operator in producing consistent images, using a tissue equivalent phantom.
- To assess the reliability of the B-mode textural features of clinical images.
- To quantify the textural features value for normal tissue, cysts, and malignant tissue.
- To determine the discriminatory ability of the textural features used in this study.

1.4 Thesis Structure

This study is divided into two phases. The first phase mainly examines the factors that may potentially influence the image texture, which will in turn affect the accuracy of the texture analysis techniques. This phase can further be divided into two parts: (i) phantom studies and (ii) assessment of clinical images. The second phase mainly determines the ability of the texture features to discriminate B-mode ovarian images. Figure 1.1 illustrates the approach used to achieve the objectives outlined in Section 1.3.

This thesis is structured into eight chapters which closely follow the flow of the methodology:

Chapter 1: This chapter gives an overview of the whole thesis.

Chapter 2: This chapter critically evaluates current techniques used to diagnose ovarian cancer and related issues. The alternative techniques are also reviewed in this chapter.

Chapter 3: This chapter provides the basic concepts of B-mode ultrasound and texture analysis. The common methodology and statistical measures employed in Chapter 4, Chapter 5, Chapter 6 and Chapter 7 are also outlined in this chapter.

Chapter 4: This chapter assesses the factors that may affect the extracted texture features.

Chapter 5: This chapter assesses the reliability of the B-mode image due to the ultrasound system as well as human factors.

Chapter 6: This chapter assesses the reliability of the B-mode clinical image and the ROI segmentation process.

Chapter 7: This chapter determines the discriminatory ability of the selected textural features.

Chapter 8: This chapter provides a summary of the whole thesis and presents the conclusion of this study.

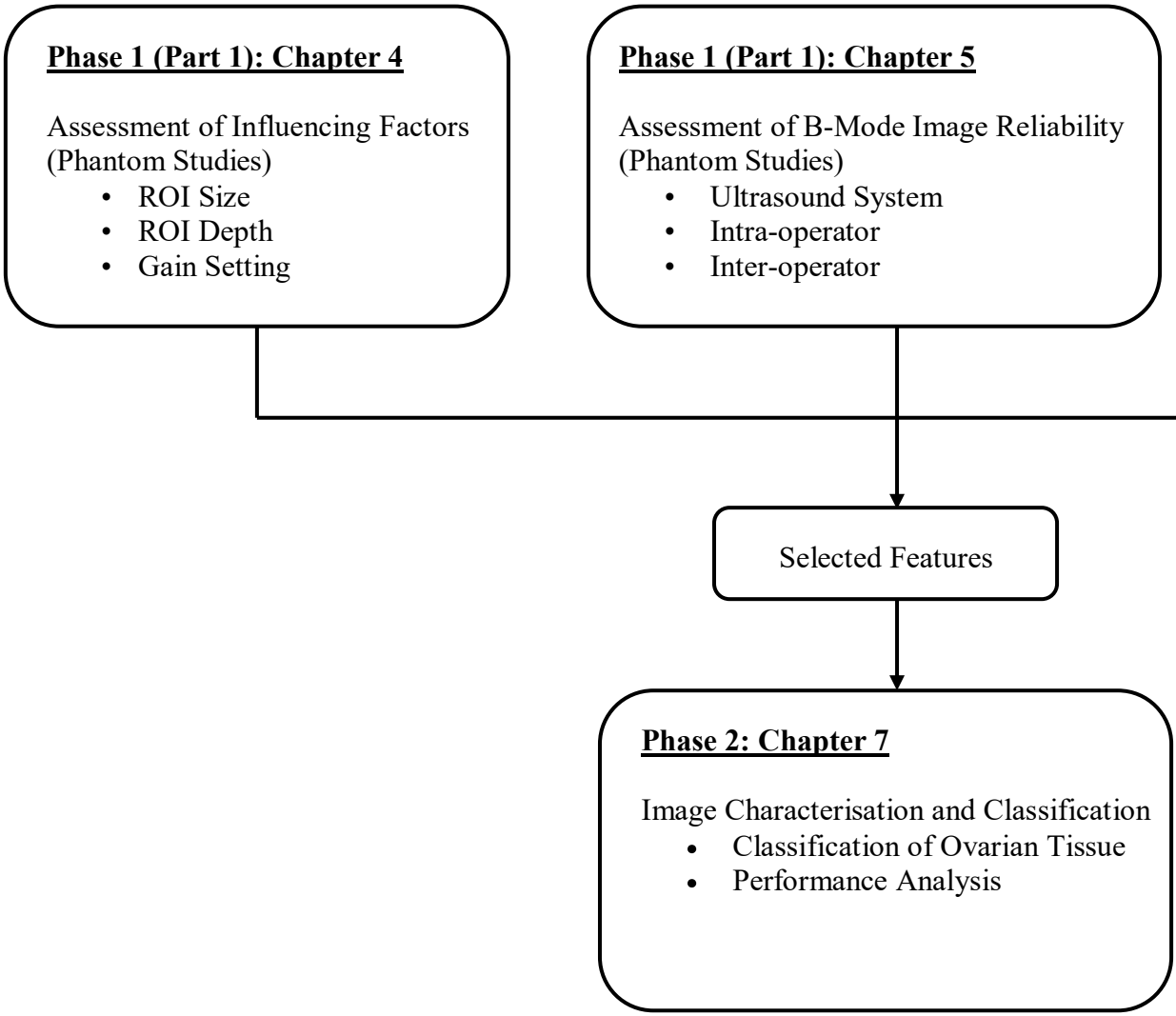


Figure 1.1: Methodology diagram.

Chapter 2
Literature Review

This chapter gives an overview of several key topics related to this thesis, and critically evaluates current techniques used to diagnose ovarian cancer and related issues. It is broken down into several sections. Section 2.1 briefly gives an introduction to ovarian cancer and quantitative image characterisation techniques. Section 2.2 provides an overview of ovarian cancer and current techniques used to diagnose ovarian cancer. Section 2.3 discusses the objective characterisation methods available, with an emphasis on texture analysis technique. Section 2.4 provides a review of the texture analysis techniques applied on medical images. Finally, section 2.5 summarises this chapter.

2.1 Background

Ovarian cancer is the fifth most common cancer in UK women and is the leading cause of death from gynaecologic malignancies (Cancer Research UK, 2010(a)). It is also the most frequent cause of death from gynaecologic malignancies in western world (Kurjak *et al.*, 2002). Early diagnosis is a crucial factor for the prognosis. To diagnose ovarian cancer, results from tumour marker analysis are normally used in conjunction with the findings from imaging modalities such as ultrasound, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT). Although ultrasound is considered the main imaging modality for ovarian cancer triage (Kinkel *et al.*, 2000), there is concern about the reproducibility of the diagnosis mainly due to the subjective nature of the interpretation of the images.

Several techniques such as radio-frequency (RF) data analysis and image texture analysis are able to quantitatively characterise the radiology data: thus, they avoid subjectivity. Of the two, image texture analysis has attracted more interest from researchers, and has been employed on various imaging modalities to characterise and discriminate normal tissue from diseased tissue of various organs.

2.2 Ovarian cancer

2.2.1 Overview of ovarian cancer

Cells normally grow and divide to form new cells. Cancer occurs due to the uncontrolled and abnormal growth and division of the cells. These extra cells will form a mass of tissue called tumour which can be benign (non-cancerous) or malignant (cancerous). Benign tumours do not invade the tissues around them and do not spread to other parts of the body (metastasis). Unlike benign tumours, malignant tumours however, have the ability to invade and destroy the tissues around them. Cancer cells are able to metastasis via the blood stream or the lymph system and spread to other parts of the body. They also able to metastasis via transcoelomic spread (peritoneal seeding across the abdominal cavity). According to Tan et al. (2006), after direct extension, transcoelomic spread is the most common route of metastasis in epithelial ovarian cancer.

Ovarian cancer refers to the cancer of the ovaries (the women's reproductive system). The ovaries are a pair of oval-shaped organs measuring 2-4 cm in diameter and are located in the pelvis, one on each side of the uterus (figure 2.1). Two distinct functions

of the ovaries are: (i) to produce the eggs, and (ii) to produce the female hormones (estrogen and progesterone).

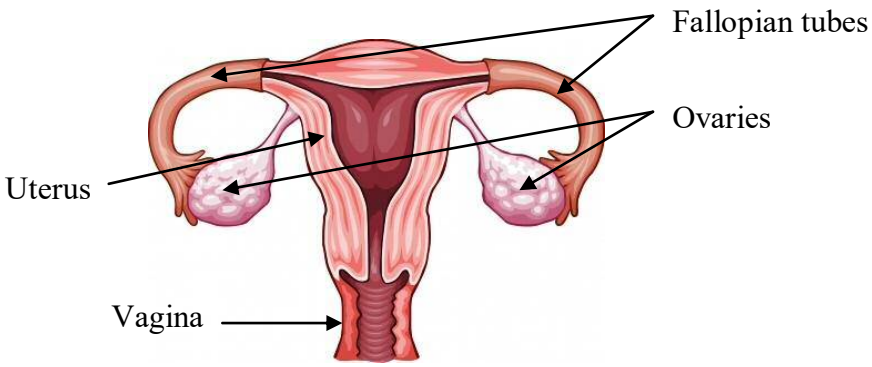


Figure 2.1: The ovaries

In general, ovarian cancer can develop at any age, but is most likely to occur in older women between the ages of 40 to 65 years (Crum, 2004). Examples of benign ovarian tumours are serous adenomas, mucinous adenomas, and Brenner tumours. Examples of malignant ovarian tumours are serous adenocarcinomas and mucinous adenocarcinomas, which originated from epithelial cells. Epithelial cell origin tumour accounts for 85% of overall ovarian cancer cases (Jeong *et al.*, 2000). According to Russell (1994), most epithelial tumours occur in peri-menopausal and post-menopausal age group with the mean age of 55 years.

Ovarian cancer is the leading cause of death from gynaecologic malignancies (Jeong *et al.* 2000; Fishman *et al.*, 2001; Chan and Selman 2006; ACS, 2007; Cancer Research UK, 2010(a)) with nearly 7,000 new cases diagnosed every year (Cancer Research UK, 2009(a)). For instance, 6,720 ovarian cancer incidences were diagnosed in the UK in 2007 (Cancer Research UK, 2010(a)). It has been reported that in 2008, around 4,370 women died of ovarian cancer in the UK, 29,000 in Europe and 140,000 worldwide (Cancer Research UK, 2010(a)).

2.2.2 Ovarian Cancer: Risk factors

Although extensive research has been carried out, the real causes of ovarian cancer are unknown. Nevertheless, the two most influential factors associated with the risk of developing ovarian cancer are the increase in age and the presence of certain gene mutations. Other factors that may affect the risks of developing ovarian cancer are briefly explained below.

- Parity - The risk is lower in women that have had children, compared to women who have no children.
- Breastfeeding – The risk is reduced in parous women who have ever breastfed, compared to those who have never breastfed.
- Infertility - Nulligravid women who have been attempting pregnancy for more than five years have an increased odds ratio compared to women who have been trying to conceive for less than a year.
- Oral contraceptives - The use of oral contraceptives reduces the risk.
- Tubal ligation - Tubal ligation has a protective effect on ovarian cancer with an estimated reduced risk of between 18% and 70%. Hysterectomy may also reduce the risk.
- Hormone replacement therapy (HRT) - Use of hormone replacement therapy (HRT) is associated with an increase in risk of ovarian cancer. The risk is not increased, however, in women using HRT for less than five years. The risk reduces once a woman stops taking HRT.
- Body mass index (BMI) - Study showed an increased risk of ovarian cancer in a mixed group of pre and post-menopausal women in relation to higher BMI.
- Family and previous cancers history - On average, women with a mother or sister diagnosed with ovarian cancer have a higher risk of developing ovarian cancer. Similarly, women with a previous breast cancer have double the risk of ovarian cancer, and the risk is almost four-fold for women diagnosed with breast cancer before the age of 40.

More detailed information on ovarian cancer risk factors is available online at: http://info.cancerresearchuk.org/cancerstats/types/ovary/riskfactors/.

2.2.3 Ovarian Cancer: Symptoms

According to Chan and Selman (2006), the symptoms of ovarian cancer are non-specific. Nevertheless, according to Flam *et al.* (1988), Goff *et al.* (2000), Olson *et al.* (2001) and Vine *et al.* (2003) as quoted by Cancer Research UK (2009(b)), the more common symptoms of ovarian cancer include abdominal pain and bloating, fatigue, weight loss, urinary symptoms, and occasionally abnormal vaginal bleeding.

Apart from that, loss of appetite, nausea, lower back pain, and shortness of breath may also be an indication of ovarian cancer. Nevertheless, it is important to note that these symptoms could also be due to other diseases.

2.2.4 Ovarian Cancer: Types, Staging and Grading

2.2.4.1 Ovarian Cancer: Types

Ovarian cancer is grouped into three major types according to the tissue of the ovary it originated from: (i) epithelial, (ii) germ cell, and (iii) sex cord-stromal cell. Epithelial tumours are the most common with 85% of cases (Jeong *et al.*, 2000). Epithelial tumours are rare before puberty. Their prevalence increases with age and peaks in the 6^{th} and 7^{th} decades of life (Jeong *et al.*, 2000). Germ cell tumours account for 5%-10% of total cases. Unlike epithelial tumours which usually affect middle aged and older women, germ cell tumours tend to be found in younger women where the peak incidence is in the early 20s. Sex cord-stromal tumours begin in the connective cells.

These connective cells hold the ovaries together and produce female hormones. Sex cord-stromal tumours can affect all age groups and account for less than 5% of the total cases.

2.2.4.2 Ovarian Cancer: Staging

Staging relates to how far the cancer has spread into other part of the body. Since it indicates the size and extent of the spread of the tumour, accurate cancer staging is an important factor in predicting the survival rate. Since cancer at different stages are managed and treated differently, staging should also help determine the appropriate management and treatment such as surgical and chemotherapeutic planning (Kurtz, 1999).

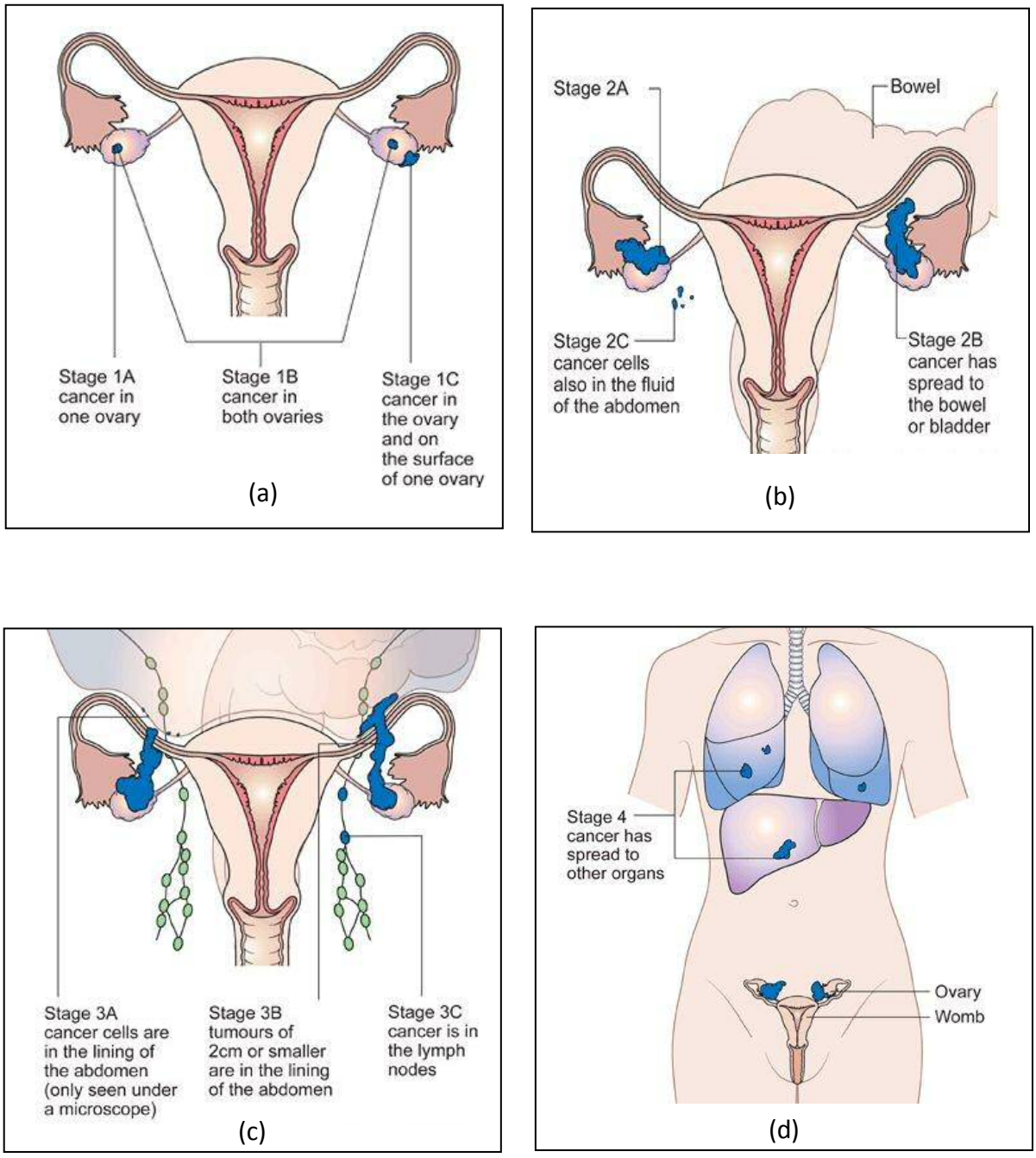
Staging is based on a 1-4 system, called the FIGO system after its originators. The document that details out the FIGO ovarian cancer staging is available online (*http://www.figo.org/files/figo-corp/docs/staging_booklet.pdf*). In brief, the four stages of ovarian cancer are as follows and the illustrations are given in figure 2.2(a) to figure 2.2(d):

Stage 1: The cancer is confined to one or both ovaries and has not yet spread to other areas.

Stage 2: The cancer can be found outside of the ovary or ovaries, but has spread no further than the pelvic region (uterus, bladder, lower intestine).

Stage 3: The cancer is limited to the peritoneal cavity.

Stage 4: The cancer has spread beyond the abdomen such as the liver, lungs and brain.



Figures 2.2(a) to 2.2(d) illustrate the extent of the spread for stage 1 to stage 4 of ovarian cancer (*adopted from: http://www.cancerhelp.org.uk/type/ovarian-cancer/treatment/stages-of-ovarian-cancer*).

2.2.4.3 Ovarian Cancer: Grading

The grade indicates how much the tumour resembles normal tissue in appearance and the likelihood of the cancer to spread. The term used is differentiation, with the well differentiated tumours appearing more like the normal tissue counterpart compared to the poorly differentiated ones. Grading is divided into four: Grade 0, grade 1, grade 2, and grade 3.

Grade 0: Tumours of low malignant potential, also known as borderline tumours are the most well-differentiated, and least aggressive.

Grade 1: Low grade well-differentiated tumours.

Grade 2: Moderately differentiated.

Grade 3: Poorly differentiated: least normal, generally the most aggressive.

More detailed information on ovarian cancer grading is available online at:

http://www.ovarian.org.uk/ovariancancer/isitovariancancer/staging.asp.

2.2.5 Ovarian cancer: Diagnosis and Treatment

Ovarian cancer is diagnosed at an advanced stage in most patients due to the non-specific signs and symptoms of this disease. Diagnosis of ovarian cancer may involve the following methods: physical examination, blood test (CA-125), abdominal fluid aspiration, imaging procedure (ultrasound / CT / MRI), laparoscopy, and laparotomy. The only way to confirm a diagnosis is to remove a tissue from the suspicious area and examine it under a microscope (ACS, 2001; Marshall, 2008). This procedure is called a biopsy.

Before ultrasound became widely available, the finding of a pelvic mass or a palpable ovary in a postmenopausal woman was considered an indication for surgery (Valentin,

1999). It is obvious that imaging of suspected ovarian malignancy requires not only a high sensitivity, but also reliable characterisation of the lesions to allow accurate and consistent diagnosis. This is particularly important since the symptoms of ovarian cancer are non-specific, and the patient management and treatment are tumour stage dependent.

2.2.5.1 Tumour Markers

The most extensively investigated ovarian cancer associated tumour marker (antigen) is CA-125 (Jeong *et al.,* 2000; Luo *et al.*, 2003). The concept of using tumour markers either for screening or diagnostic tests of ovarian cancer is dependent upon identifying an abnormal level of a particular marker in serum, reflecting a systemic effect of disease in the ovary. According to Agency for Healthcare Research and Quality (AHRQ) report (2006), as quoted by The Royal College of Obstetricians and Gynaecologists (RCOG) in their document "*Guideline No: 34 Ovarian Cysts in Postmenopausal Women*", elevated levels of CA-125 are detected in approximately 80% of ovarian cancer at the time of diagnosis. CA-125 however, is raised in only 50% of stage 1 cases (Skates *et al.*, 2000; RCOG, 2003). It is important to note, elevated CA-125 serum levels have also been reported in a variety of benign conditions which could potentially affect the specificity, particularly for diagnosis of early-stage disease (Luo *et al.*, 2003).

According to Chan and Selman (2006), the use of CA-125 alone for early diagnosis is unhelpful, where tests carried out by Moss *et al.* (2005) discovered that only 20% (39 out of 195) of the abnormal results were due to ovarian cancer. Nevertheless, CA-125 remains the most useful tumour marker clinically (Chan and Selman, 2006) and has been recommended by RCOG to be used together with transvaginal B-mode ultrasound

for the assessment of ovarian cysts (RCOG, 2003). It is the most widely used tumour marker for monitoring of disease to determine if second-look surgery is required (Skates et. al, 2000).

Other tumour markers include carcinoembryonic antigen (CEA) for epithelial tumours; alpha-fetoprotein (AFP) and beta-human chorionic antigen (β-HCG) for germ cell tumours; and inhibin B for granulosa cell tumours (Turkington and Edelson, 2005).

2.2.5.2 Ultrasound-based Ovarian Cancer Diagnosis

Ultrasound has been used in medical imaging for over half a century (Hangiandreau, 2003), and is commonly regarded as the preferred imaging modality in the study of the female pelvis (Derchi *et al.*, 2001). Improvements in technology have been followed by widespread acceptance and use of ultrasound in medical diagnosis. The introduction of transvaginal transducer for example, has improved visualisation of normal ovarian function and ovarian tumours (Twickler and Moschos, 2010). According to Halliwell (2010), currently, ultrasound is responsible for 20% of all diagnostic images. Applications have progressed from simple measurements of anatomical dimensions to detailed screening for fetal abnormalities, detection of subtle changes in tissue texture, and detailed study of blood flow in arteries (Martin, 2010). Ultrasound is currently one of the most significant, widely used, and versatile imaging modalities in medicine (Hangiandreau, 2003).

Analysis of ultrasound images can assist in differentiating between benign and malignant lesions. It is the imaging modality of choice in the evaluation of suspected adnexal masses (Togashi, 2003) as well as ovarian masses (Kinkel *et al.*, 2000). According to Kinkel *et al.* (2000), ultrasound is the main triage method for ovarian cancer prior to treatment. Ultrasound-based assessment for ovarian mass characterisations can be summarised under the following techniques: (i) morphologic information which is based on B-mode images, and (ii) blood flow information which is based on Doppler imaging.

A wide range of performance of ultrasound examination in ovarian cancer diagnosis has been reported. Based on their review, Kinkel *et al.* (2000) reported that accuracy of

ultrasound is 65%–94% for B-mode ultrasound, 35%–88% for colour Doppler flow imaging, and 48%–99% for Doppler arterial resistance measurements. The Royal College of Obstetricians and Gynaecologists (RCOG), however, in its document "*Guideline No: 34 Ovarian Cysts in Postmenopausal Women*" quotes DePriest *et al.* (1994) for ultrasound performance as having sensitivity of 89% and specificity of 73% when using a morphology index.

Despite vast studies and reviews being carried by many researchers, which ultrasound technique and diagnostic criteria provide the best ovarian lesion characterisation, however, remains inconclusive (Kinkel *et al.*, 2000). Nevertheless, ultrasound techniques that combine gray-scale ultrasound morphologic assessments with tumour vascularity imaging information (colour Doppler flow imaging) in a diagnostic system, are significantly better in ovarian lesion characterisation compared to using Doppler arterial resistance measurements, colour Doppler flow imaging, or gray-scale ultrasound morphologic alone (Kinkel *et al.*, 2000; Togashi, 2003).

(i) *B-mode (Brightness-mode) Ultrasound*

The diagnosis of ovarian cancer from B-mode ultrasound is based on morphological features. The features include the cystic and solid tumour composition, the presence and type of septations, and papillarities. According to Aletti *et al.* (2007), features highly suggestive of ovarian cancer include the presence of a complex ovarian mass, with both solid and cystic components, sometimes with internal echoes and/or septations. The large size of an ovarian mass also has been found to be a significant factor in predicting ovarian cancer (Twickler and Moschos, 2010). Figure 2.3(a) and 2.3(b) are examples of B-mode images of an ovarian tumour.

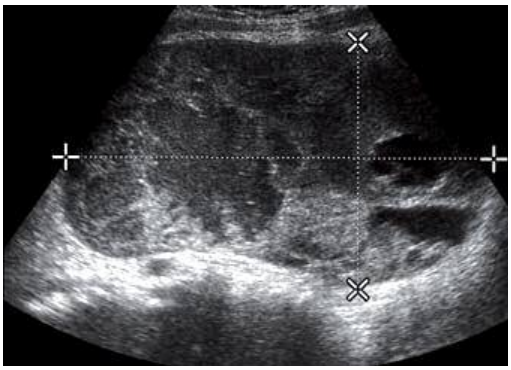


Figure 2.3 (a) Very large, complex solid-cystic mass (*adopted from Twickler and Moschos, 2010*).

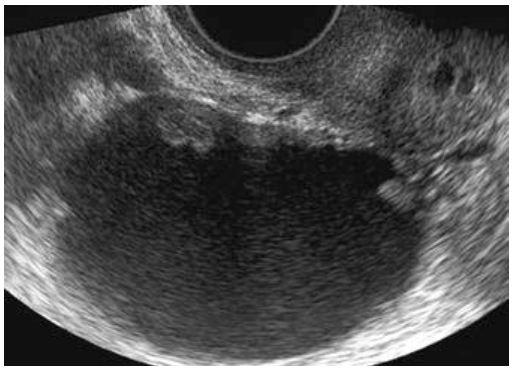


Figure 2.3 (b) Cystic mass with internal echoes and multiple papillaries (*adopted from Twickler and Moschos, 2010*).

Although evaluation of ultrasound findings is an excellent method for discriminating between benign and malignant adnexal tumours (Ameye *et al.*, 2009), it is highly dependent on the operator's experience (Wang *et al.*, 2002; Shung, 2006; Huang *et al.*, 2008; Rocha *et al.*, 2011). Hence, the lack of standardised terms and procedures in the interpretation of gynaecological ultrasound is a general cause of concern (Timmerman, 2000). In order to provide a more standardised terminology and procedure, a group of researchers from the International Ovarian Tumor Analysis (IOTA) group has produced a document entitled "*Terms, Definitions and Measurements to Describe The Sonographic Features of Adnexal Tumors: A Consensus Opinion From The International Ovarian Tumor Analysis (IOTA) Group*" (Timmerman *et al.*, 2000).

Several scoring systems for morphological features have also been introduced to increase the test performance. Table 2.1(a) – (c) summarise the criteria used for the scoring system published by several groups.

Table 2.1(a): Detailed description of ultrasound scoring systems by Sassone *et al.*, 1991.

Criteria	**Score**				
Morphology	1	2	3	4	5
Inner wall structure	Smooth	Irregularities ≤ 3 mm	Papillarities >3 mm	Not applicable, mostly solid	-
Wall thickness (mm)	Thin (≤ 3)	Thick (> 3)	Not applicable, mostly solid	-	-
Septa (mm)	None	Thin (≤ 3)	Thick (> 3)	-	-
Echogenicity	Sonolucent	Low echogenicity	Low echogenicity with echogenic core	Mixed echogenicity	High echogenicity
	Minimum total score for suggestive of malignancy: > 9				

Table 2.1(b): Detailed description of ultrasound scoring systems by DePriest *et al.*, 1993.

Criteria	**Score**				
Morphology	1	2	3	4	5
Cystic wall Structure (thickness)	Smooth (< 3 mm)	Smooth (≥ 3 mm)	Papillary projection (< 3 mm)	Papillary projection (≥ 3 mm)	Predominately solid
Volume (cm3)	<10	10-50	>50 – 200	>200-500	>500
Septum structure	No septa	Thin septa (< 3 mm)	Thick septa (3 mm to 1 cm)	Solid area (≥ 1 cm)	Predominately Solid
	Minimum total score for suggestive of malignancy: ≥ 5				

Table 2.1(c): Detailed description of ultrasound scoring systems by Lerner *et al.*, 1994.

Criteria	**Score**				
Morphology	1	2	3	4	5
Wall structure	Smooth or small irregularities < 3 mm	-	Solid or not applicable	Papillarities ≥ 3 mm	-
Shadowing	Yes	No	-	-	-
Septa	None or thin (< 3 mm)	Thick (≥ 3 mm)	-	-	-
Echogenicity	Sonolucent or low-level echo or echogenic core	-	-	Mixed or high	-
	Minimum total score for suggestive of malignancy: ≥ 3				

Although the use of a scoring system helps to improve the test performance, according to Gramellini *et al.* (2008), the existence of multiple scoring systems may pose problems in clinical practice mainly for two reasons: (i) there are so many of them and, (ii) the parameters to be considered are often considerably complex. Moreover, there is a lack of consistency because of the variables used and their relative importance was arbitrarily defined (Timmerman *et al.*, 2008). They quoted an example that there was no general agreement as to when a solid protrusion from a cyst wall projects sufficiently into the cyst cavity for it to be termed a solid papillary projection.

Risk of malignancy index (RMI)

The risk of malignancy index (RMI) has been developed to guide clinicians to identify those patients with pelvic masses who are likely to have ovarian cancer. It is an effective way of triaging women into those who are at low, moderate, or high risk of malignancy (RCOG, 2003). To date, various versions of the risk of malignancy index have been published. According to Chan and Selman (2006), the most commonly used version in the United Kingdom is that published by the Royal College of Obstetricians and Gynaecologists in October 2003 (RCOG, 2003).

The formula for calculating the risk is:

$$RMI = U \times M \times CA\text{-}125 \qquad \text{Equation (1)}$$

where U is the ultrasound score and M is the menopausal status.

One point is scored for each of the following ultrasound characteristics: multi-loculated cyst, evidence of solid areas, evidence of metastases, presence of ascites, and bilateral lesions. U = 0 (for ultrasound score of 0); U = 1 (for ultrasound score of 1); U = 3 (for ultrasound score of 2-5); and M = 3 for all postmenopausal women.

The risk is considered high (75%) if the RMI exceeds 250. It is worth noting that, one of the issues of RMI is that, it is dependent on serum CA-125, which is a non-specific marker that leads to unnecessary surgical intervention (Ortashi *et al.*, 2008).

In an attempt to improve the diagnostic accuracy, a team at the Doppler Ultrasound Department in the University Hospital of Wales has introduced a scoring system named Cardiff Malignancy Index (CMI). CMI is independent of CA-125. Instead, it combines

the finding of B-mode ultrasound with Doppler ultrasound. Even though the initial results showed good sensitivity and specificity, Ortashi *et al.* (2008) stated that larger numbers and inter-observer studies were required to validate their findings.

(ii) Doppler Ultrasound

The Doppler effect enables ultrasound to be used to detect the motion of blood (Horskins, 2010). Doppler ultrasound has been available to clinicians for nearly 40 years (Boote, 2003). It has been used mainly for discriminating between benign and malignant lesions (Alcazar, 2006). Doppler ultrasound is used to measure blood velocity by means of the Doppler frequency shift of the echoes received from red blood cells (Rubin, 1994) and allows the assessment of tumour vascularity. The basis is that, malignant neoplasms have active blood vessel creation (angiogenesis) compared to normal or benign neoplasms.

In colour Doppler imaging technique, the flow detection within a mass is an indication of malignancy (Kinkel *et al.*, 2000). While acknowledging that the amount and appearance of intra-tumoral blood vessels give clues to the nature of the tumour, Timmerman (2000) raised two issues about the colour Doppler ultrasound: (i) the assessments are subjective, and (ii) the assessments depend on the quality of the equipment and the settings.

In Doppler arterial resistance technique (Doppler signals analysis), a threshold value is used to characterise the mass. Parameters such as pulsatility index (PI), resistive index (RI), and peak systolic velocity (PSV) have been used. In 15 data sets reviewed by Kinkel *et al.* (2000), in which a PI was used, a threshold value of 1.0 was used in 67%

of the studies; other threshold values were 0.62-1.50. For the RI, out of 16 data sets reviewed, the authors reported that a large range of threshold values were used: 0.4 - 0.8.

The major limitation of RI, PI, and PSV is that, the range observed in malignant masses overlaps with that observed in benign masses (Kurjak *et al.*, 2003; AHRQ, 2006), making this approach impractical from clinical point of view (Valentin *et al.*, 1994). This argument is supported by Laban *et al.* (2007) when the authors concluded that RI measurement cannot be used alone for the detection of malignant ovarian tumours, because there is too much overlap between RI of benign and malignant ovarian tumours. The overlap in the range value between benign and malignant masses limits the effectiveness of the application of threshold value; that is what cut-off value is to be used. Consequently, according to Kurjak and Kupesic (2003), the attitudes concerning the usefulness of Doppler ultrasound for the assessment of ovarian vascularity in the detection of adnexal malignancies have been equally divided.

A new technique of Doppler ultrasound provides three-dimensional (3D) imaging. Three-dimensional (3D) ultrasound was first demonstrated in 1970s while the first commercial 3D ultrasound became available in 1989 (Prager *et al.*, 2010). Three-dimensional ultrasound exploits the real-time capability of ultrasound to build a volume that can then be explored using increasingly affordable high-performance workstations (Nelson, 2006).

According to Benacerraf (2008), to date, there is not enough information available to determine whether 3D imaging of the adnexa adds significant information that is not available from standard two-dimensional (2D) scanning. Even though there are studies reporting that 3D ultrasound may be useful for differentiating benign from malignant

ovarian tumours (Kurjak *et al.*, 2002; Alcazar and Castillo, 2005; Testa *et al.*, 2005), other studies either suggest the need for further research (Rieck *et al.*, 2006; Alcazar, 2006) or is to be used as an adjunct to morphologic assessment (Wilson *et al.*, 2006). While Fishman *et al.* (2001) states that the clinical value of 3D ultrasound for the early detection of ovarian carcinoma has yet to be determined, Dai *et al.* (2008) concluded that it did not improve the diagnostic accuracy for the prediction of malignancy in adnexal masses and further emphasise that 2-D transvaginal sonography may still remain an important modality for the prediction of adnexal malignancy. In addition to that, Jokubkiene *et al.* (2007) concluded that objective quantification of the colour content of the tumour scan using 3D ultrasound did not seem to add more to B-mode imaging, compared to the subjective quantification using 2D power Doppler ultrasound. In a recent review by Alcazar and Jurado (2011), they concluded that further studies are needed to establish the role of 3D ultrasound in clinical practice in gynaecological oncology.

2.2.5.3 Other Imaging Modalities

Computed tomography (CT) and magnetic resonance imaging (MRI) may be used to help in the diagnosis at presentation, to assess the extent of disease, to assess the residual disease, to assess the response to treatment, and to assess the recurrence of the disease (Jeong *et al.*, 2000; Sohaib and Reznek, 2002).

Computed tomography (CT) has been used primarily in patients with ovarian malignancies, either to assess disease extent prior to surgery, or as a substitute for second-look laparotomy (Jeong *et al.*, 2000). Computed tomography (CT) imaging of the abdomen or pelvis allows comprehensive evaluation of all potential sites of peritoneal implants or lymphadenopathy as well as of the primary tumour site. According to Kurtz *et al.* (1999), CT was superior in the diagnosis of malignancy; however, ultrasound performed equally well with CT in detection of masses. Compared to ultrasound, besides being more expensive, another major disadvantage of CT imaging is that, it involves ionising radiation (Togashi, 2003).

Comparison studies of ultrasound and MRI have shown MRI to be better in differentiation of benign from malignant masses (Kurtz *et al.*, 1999). The accuracy of MR imaging in the diagnosis of mature cystic teratomas, endometriomas, and leiomyomas is well established, particularly due to its superb contrast resolution (Jeong *et al.*, 2000). Agency for Healthcare Research and Quality (AHRQ) however, in its report on *Management of Adnexal Mass* (2006) stated that, in pooled analysis, the combination of ultrasound morphology and Doppler blood flow is comparable to MRI. Nevertheless, the major limitations of MRI are its limited availability and long scanning times (Forstner *et al.*, 1995). Jeong *et al.* (2000) suggests that MRI is better reserved for

problem solving when ultrasound findings are non-diagnostic. The reason is that, although MRI is more accurate for diagnosis, it is also more expensive.

It is important to note that, compared to the CT and MRI, ultrasound is relatively in-expensive, portable, and widely available. Furthermore, ultrasound does not cause any known risk to the patient. According to RCOG (2003), there is no routine role yet for Doppler, MRI, CT or PET, because the relative expense and limited availability of these modalities, which can cause the delay in the referral and treatment.

2.2.5.4 Laparoscopy and Laparotomy

The definitive diagnosis of ovarian cancer is by performing surgery (laparoscopy or laparotomy) and biopsy (removal of tissue for examination under a microscope). Laparoscopy is the insertion of a thin, lighted tube (called a laparoscope) through the abdominal wall to inspect the inside of the abdomen and remove tissue samples, while laparotomy is a surgical incision made in the wall of the abdomen (NCI, 2006).The choice of surgery depends on the probability of malignancy (Kinkel *et al.*, 2000).

Laparotomy is performed if the suspicion is strong for malignancy (Benedet *et al.*, 2000). According to Jeong *et al.* (2000), since the delay in performing definitive staging laparotomy may be associated with a worse prognosis, laparoscopy on the other hand, is reserved for patients with masses that are non-suspicious based on imaging findings.

Clinical judgment is a major factor in surgical decision-making and pertinent in the approach to a pelvic mass in the young woman of reproductive age (Benedet *et al.*, 2000). In many young patients with non-malignant ovarian lesions such as endometriosis and benign cysts, treatment with laparoscopy can avoid laparotomy procedure (Jeong *et al.*, 2000).

In a postmenopausal woman who is not suitable for conservative management, RCOG has recommended an oophorectomy procedure even when the risk of malignancy is low (RCOG, 2003). Exploratory laparotomy is necessary in all cases of suspected ovarian cancer to confirm the diagnosis, to determine the extent of the disease (by staging), and to resect the tumour (Jeong *et al.*, 2000).

2.2.6 Ovarian Cancer: Prognosis

The most important determinant of survival for ovarian cancer is the stage (extent of spread) of disease at diagnosis (Cancer Research UK, 2010(c)). The prognosis is considerably changed by the extent of spread. According to Kurtz *et al.* (1999), the 5-year survival rate is 85% if the cancer is confined to the ovaries (stage 1), 55% if the cancer has spread into pelvis (stage 2), 14% if the cancer has spread to the abdominal (stage 3), and 4% if the cancer has spread beyond the abdomen (stage 4). However, there was a slight difference in the survival rate recorded in the Thames Cancer Registry area, from 1992 to 1996 as published by Cancer Research UK (Cancer Research UK , 2010(c)) in their website (Table 2.2):

Table 2.2: Five-year relative survival for ovarian cancer patients diagnosed in the Thames Cancer Registry Area, 1992-1996

Stage	Proportion of women diagnosed	Five- year relative survival	95% CI
Local	20%	73%	70.6-75.9
Direct extension	8%	34%	30.7-38.2
Regional lymph nodes involved	1%	27%	16.4-37.8
Distant metastases	40%	16%	14.7-17.8
Stage unknown	31%	39%	36.2-42.1

(*adopted from http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/*)

2.3 Objective Characterisation of Medical Images

Over the last few decades, a considerable number of research studies have applied a variety of quantitative analysis methods in order to improve the differentiation of tissues. Because each individual perceives grey-scale images differently, it leads to the variability in image interpretation (Pierson and Adams, 1995). Quantitative analysis methods on the other hand, are able to eliminate the subjective interpretation of the image (Jaqaman and Danuser, 2009).

There are two quantitative approaches that have been explored extensively: (i) analysis of the radio-frequency (RF) data, and (ii) texture analysis of the greyscale image. Texture analysis, however, gains more attention from researchers for several reasons:

(i) Texture analysis is generally easier to implement compared to using the RF data (Morris, 1988(b)).
(ii) RF data analysis requires specialised equipment not normally available in clinical settings (Mass, 1994).
(iii) Equipment to obtain the RF data from the scanner is not readily available in most ultrasound centres (Coleman, 2007).
(iv) Commercial ultrasound typically provides B-mode images only and not the RF data (Noble, 2010).
(v) The RF data must be sampled at a very high rate and requires a large storage space (Alqahtani, 2010).

2.3.1 Analysis of RF Data

The use of RF ultrasonic data for objective quantification started in as early as late 70s when Lerski *et al.* (1979) applied this technique to study alcohol induced diffuse liver disease. In 1981, Lerski *et al.* extended their work by differentiating between normal

and several diffuse alcoholic liver disease aetiologies: alcoholic liver disease, primary biliary cirrhosis, haemochromatosis and chronic active hepatitis. Both studies gave encouraging results. In 1980, Roman used RF data (ultrasound attenuation coefficient) to characterise liver pathology while in 1995, Chan analysed RF data to quantitatively characterise plaque, blood and vessel wall of carotid atheroma, and reported a substantial improvement in the characterisation.

Recently, Advanced Medical Diagnostics (AMD) has developed a technique called *HistoScanning*, to quantitatively characterise prostate tissue based on the analysis of RF data (AMD, 2010(a)). On the assessment of this technique which involved 29 men, Braeckman *et al.* (2007) concluded that this technique has the potential to identify and characterise prostate cancer. In their further study of this technique, on relatively small prostate cancer lesions, the authors reported that their results showed that the lesions can be accurately detected, but stated that verification in a larger group was needed (Braeckman *et al.*, 2008).

Using the same technique developed by AMD, Spethmann *et al.* (2010) carried out a study to predict the chance of neuro-vascular bundle (NVB) preservation in prostate cancer and reported that the results encouraged further investigation. With regards to the use of this technique on ovary, although Lucidarme *et al.* (2010) reported that the analysis of RF data from ultrasound using *HistoScanning* technique is highly sensitive for the diagnosis of malignant ovarian masses, the company (AMD) deems its application on ovarian cancer as still in the development stage (AMD, 2010(b)).

2.3.2 Image Texture Analysis

Image texture is defined as a function of the spatial variation in pixel intensities (Tuceryan and Jain, 1998) while the analysis of image texture refers to the branch of imaging science that describes the image properties by textural features (Nailon, 2010). Image texture may carry substantial information about the structure of physical objects. Since the output of texture analysis is numerical, the ambiguity in recognition as well as discrimination of the lesion may be avoided (Chan and McCarty, 1990). According to Tourassi (1999), image texture analysis is currently undergoing great development and utilization within the field of medical imaging. Texture analysis concepts and methods are described later in this thesis (Chapter 3).

Texture analysis has been utilized in a variety of application domains such as surface inspection, document processing, and medical imaging. It has played a major role in some mature domains, such as remote sensing (Tuceryan and Jain, 1998). In the medical domain, although interesting results were reported, early attempts of texture analysis application were not successful due to several main obstacles such as computers were not sufficiently powerful (Doi, 2007; Giger *et al.*, 2008), advanced image-processing techniques were not available (Doi, 2007), limited quality of the image (Giger *et al.*, 2008), and digital images were not easily accessible (Doi, 2007).

Several factors, namely the development of physical sciences, engineering, computer technology and image processing, however, have collectively contributed to further development of texture analysis. For instance, the developments in transducer design for ultrasound equipment have resulted in transducers with greater band width (Averkiou *et al.*, 1997). The broadband transducers improve the axial resolution of the

images (Browne *et al.*, 2004). Since texture analysis deals with the pixel intensity and its relationships with neighbouring pixels, resolution is one of the factors that may influence the performance. In addition, the advancement in engineering has improved the contrast and reduced the image noise (Bridal *et al.*, 2003).

The potential use of computers for analysis of medical images has been considered since 1950s with a tremendous growth over the past 20 years (Doi *et al.*, 1999; Giger *et al.*, 2008). This can be seen in a significant increase of publication in the literature. For example, in Medical Physics journal alone, more than 20 articles were published in 2008 compared to only 5 articles published in 1990 (Giger *et al.*, 2008).

Early studies utilising texture analysis used first-order statistics, such as the grey level mean and variance (Chan, 1998). Although these parameters have the advantage that they can be calculated quickly, they are unsophisticated assessors of textural pattern. According to Thijssen (1992), as computational speed of computer improved, many ultrasound tissue characterisation studies applied second-order statistics, such as the grey level co-occurrence matrix (GLCM) and grey level run length matrix (RLM).

2.4 Application of Texture Analysis on Medical Images

Texture analysis was first used for satellite image analysis (Herlidou *et al.*, 1999) and is increasingly used to evaluate the texture properties in order to improve interpretation of medical images (Dhawan, 2003). Image texture of medical images describes internal structure of human tissues or organs (Szczypinski *et al.*, 2009) as well as pathological changes (Mayerhoefer et al, 2008; Xian, 2010). Many studies have demonstrated the value of texture analysis in medical applications. The choice of texture analysis rests on its ability to quantitatively describe the image.

According to Doi (2007), early studies on texture analysis techniques on medical images were reported in the 1960s. It has been applied on various types of tissues and organs. For instance, it has been applied on carotid artery (Coleman *et al.*, 2005), breast (Alacam *et al.*, 2003), heart (Vince *et al.*, 2000), prostate (Basset *et al.*, 1993), pancreas (Lefebvre *et al.*, 1998), and liver (Smutek *et al.*, 2006).

Texture analysis has also been applied across a various types of imaging modalities such as ultrasound, magnetic resonance imaging (MRI), computer tomography (CT), conventional x-ray, mammogram, optical coherence tomography (OCT) and capsule endoscopy (CE). According to Tsai and Kojima (2005), the employment of texture analysis in medical imaging has been proved to be valuable, particularly for MRI, CT, and ultrasound. Most of the previous works in texture analysis involves MRI images because of the great amount of detail provided by this technique (Castellano *et al.*, 2004). Nevertheless, texture analysis of all sorts of images is possible and has been performed.

The following sections provide review of texture analysis applications on medical images and are summarised according to the type of the imaging modalities: ultrasound, conventional x-ray, CT, MRI and mammography.

2.4.1 Texture Analysis on Ultrasound Images

Many studies have been carried out to quantitatively characterise the echo-texture of B-Mode images. Texture analysis of ultrasound images lies on the principles that, if disease processes affect the structure of the tissue, the tissue should reflect an altered ultrasound signal, which will in turn give in texture features value different to the normal tissue (Morris, 1988(b)). The transformation of cancerous tissue for example, will result in the changes in the tissue characteristics such as density and elasticity. Based on this principle, it is expected that textural features derived from cancerous tissue and normal tissue will differ.

Morris (1988(b)) demonstrated that grey level co-occurrence matrix (GLCM) features derived from ultrasound images have a significant difference between the placenta of smoker and non-smokers. Similarly, using GLCM feature, Tsai and Kojima (2005) extracted textural features from ultrasound images to classify heart disease. From their findings, the authors concluded that texture analysis of ultrasound images has the potential to become clinically useful for computer-aided diagnosis of cardiomyopathy.

Texture analysis of endoscopic ultrasound (EUS) images has also received much attention. In a recent study by Das *et al.* (2008), they reported that their initial attempt to characterise pancreatic cancer from EUS images was encouraging. The authors

discovered that texture analysis was able to differentiate pancreatic cancer from chronic inflammation and normal tissue, and reported the area under the ROC curve of 0.93. Texture analysis of EUS images has previously been applied by Norton *et al.* (2001) to differentiate between pancreatic malignancy and pancreatitis; and by Loren *et al.* (2002) for analysis of lymph node metastasis in patients with oesophageal carcinoma. Both studies reported promising results.

Texture analysis has also been applied extensively on liver related diseases. In one of the most recent studies, Xian (2010) reported that textural information extracted from B-Mode sonography is an efficient feature for distinguishing liver diseases. This is supported by many promising previously reported results. For instance, Fellingham and Sommer (1984) uses texture analysis to characterise the in vivo tissue structure of the human liver and spleen. Promising results also have been reported by Kadah *et al.* (1996) in their work to quantitatively characterise the tissue of diffuse liver disease. Later, based on their experiment findings, Mojsilovic *et al.* (1998) suggested that wavelet-transform features could be a reliable method for a texture characterisation of B-mode liver images. In a more recent study, Yeh *et al.* (2003) used GLCM and wavelet features to classify six grades of liver fibrosis. Their results revealed that the best classification accuracy of two, three, four and six classes were 91%, 85%, 81% and 72%, respectively. Texture analysis also has been implemented on liver related diseases by many other researchers (Layer *et al.*, 1990; Wang *et al.*, 2002; Lee *et al.*, 2003; Ribeiro and Sanches, 2009).

Besides liver, texture analysis of ultrasound images has also been applied on thyroid gland. A promising result was reported by Smutek *et al.*(2003): their results showed that

chronic inflamed thyroid tissue could be differentiated from healthy thyroid gland tissue and reported a classification success rate of 100%.

Texture analysis technique has also been extensively applied on ultrasound images to characterise carotid plaques (Elatrozy *et al.*, 1998; Rakebrandt *et al.*, 2000; Christodoulou *et al.*, 2003; Coleman *et al.*, 2005; Stoitsis *et al*., 2006(a); Kakkos *et al.*, 2007), breast (Murmis *et al.*, 1988; Lefebvre *et al.*, 2000; Huber *et al.*, 2000; Alvarenga *et al.*, 2007; Sivaramakrishna *et al*., 2002; Liu *et al.*, 2010) , and prostate (Huynen *et al.*, 1994; Sheppard and Shih, 2005; Han *et al.*, 2008; Llobet *et al.*, 2007).

2.4.2 Texture Analysis on Other Imaging Modalities Images

Besides ultrasound, the application of texture analysis also has been studied across the various imaging modalities such as X-ray, CT, MRI, and optical coherence tomography, with promising results reported. The following sections provide brief review of texture analysis application on these modalities.

2.4.2.1 Texture Analysis on Conventional X-Ray Images

One of the early texture analysis applications on X-ray images was reported by Chien and Fu on chest X-ray images in 1974. Van Ginneken *et al.* (2002) also applied this technique on chest radiograph to produce abnormality score for the images collected from a TB mass chest screening program. Besides chest images, texture analysis also has been widely applied on bone tissue. For instance, Bocchi *et al.* (1997) revealed that their experimental results indicate a good performance of texture analysis in characterisation of bone tissue (between normal and altered tissue) with accuracy and

sensitivity of over 90%, and specificity of 81%. Similarly, based on X-ray image, Materka *et al.* (1999) concluded that texture analysis indicates potential usefulness as an aid to the diagnosis of skeletal diseases. In a more recent study, Guggenbuhl *et al.* (2006) reported that, X-ray texture analysis appeared to be a suitable method for bone micro-architecture assessment.

The application of texture analysis techniques on X-ray images is not only confined to human but also to animal images. These techniques have been used for the characterisation of bovine ovarian follicles (Singh *et al.*, 1998), for the study of the endometrial texture of cows (Schmauder *et al.*, 2008), and for the examination of cyclical changes (Herzog *et al.*, 2008) in bovines luteal.

2.4.2.2 Texture Analysis on CT Images

Texture analysis has been demonstrated to be useful in improving the interpretation of CT images. Many texture analysis studies have been applied on two-dimensional (2D) CT images. For instance, Gletsos *et al.* (2003) uses GLCM features to characterize four groups of liver tissue: normal liver, hepatic cysts, haemangioma, and hepatocellular carcinomas. In 2006, the feasibility of this technique on CT images was investigated by Xu *et al.* for the detection of interstitial lung diseases. They reported the sensitivity was between 73%-93% and the specificity was between 90%-98%. In a recently published study carried out by Kumar and Moni (2010) to characterise the liver tumour by CT, their experiment reveals encouraging results. They have suggested that the technique used could be extended for diagnosis of other types of liver diseases. Chen *et al.* (2010), using texture analysis features as an input, reported a promising intracranial

pressure (ICP) prediction model. They also suggested that their approach could potentially be extended for analysis of other body regions.

Apart from 2D CT images, texture analysis has also been applied on 3D CT images. For instance, Tesar *et al.* (2008) investigated three-dimensional CT scans of the abdominal area. Similar to Chen *et al.* (2010), Tesar *et al.* (2008) suggested the potential application of texture analysis to be extended to the other organs.

2.4.2.3 Texture Analysis on MRI Images

According to Herlidou *et al.* (2003), texture analysis was first implemented on MRI images at the beginning of the eighties. Its applications in medical fields have slowly increased over the last ten years with a large number of publications now in the literature (Lerski, 2006). Similar to the previously reviewed texture analysis applications, findings from a review of the texture analysis on MRI images are also encouraging.

Texture analysis on MRI images has been applied on multiple sclerosis (MS) related studies. For instance, in discriminating MS from cerebral microangiopathy (CM) lesions, Theocharakis *et al.* (2009) reported the best overall accuracy (88.46%) was attained. In a recent study on MS related disease, Harrison *et al.* (2010), on the classification of white matter and multiple sclerosis lesions, reported that texture parameters provide an excellent distinction: an accuracy between 96%-100% was achieved.

In another recent study on MRI images, Holli *et al.* (2010) reported significant changes in texture parameters of cerebral tissue between hemispheres and corpus callosum (CC) segments in traumatic brain injury (TBI) patients. The authors suggested that this technique may serve as a novel additional tool for detecting the conventionally invisible changes in cerebral tissue in mild TBI, which may help clinicians to make an early diagnosis.

2.4.2.4 Texture Analysis on Mammography Images

Texture analysis has also been applied on mammography-based images. For example, in 1992, Miller and Astley applied texture analysis on mammography images to discriminate between glandular and fatty regions. Out of 40 images used, they achieved a correct classification rate of 80%. A higher accuracy rate (89%) was reported by Karahaliou *et al.* (2007) when they investigate the feasibility of texture analysis in discriminating malignant from benign breast tissue.

Besides ultrasound, conventional x-ray, CT, MRI, and mammography, texture analysis techniques have also been implemented on other imaging modalities such as on optical coherence tomography (Gossage *et al.*, 2003; Baroni *et al.*, 2007) and on capsule endoscopy (CE) images (Li and Meng, 2009).

2.5 Summary

Ovarian cancer is one of the most common cancers in women and is the leading cause of death from gynaecologic malignancies. Symptoms of ovarian cancer include abdominal pain and bloating, fatigue, weight loss, urinary symptoms, and occasionally abnormal vaginal bleeding. Two most influential factors associated with the risk of developing ovarian cancer are the increase in age and the presence of certain gene mutations.

Ovarian cancer is grouped according to the area of the ovary it originated from. Staging which relates to how far the cancer has spread into other part of the body is an important factor in predicting the survival rate. Ovarian cancer staging is based on a system called the FIGO system. Another important characteristic of the cancer is the grade. It indicates how much the tumour resembles normal tissue in appearance and the likelihood of the cancer to spread.

To diagnose ovarian cancer, results from tumour marker analysis are normally used in conjunction with the findings from imaging modalities such as ultrasound, MRI, and CT. Among these three modalities, ultrasound is considered the main modality for ovarian cancer triage. B-mode ultrasound is the main ultrasound technique for ovarian cancer diagnosis. There is no routine role yet for Doppler ultrasound, MRI, or CT.

The interpretation of the ultrasound image however, is highly dependent on operator's experience, and several scoring systems for morphological features have been introduced to increase the test performance. The risk of malignancy index (RMI),

which incorporates the results from the scoring system, has been developed to guide clinicians to identify the likelihood of having ovarian cancer.

Several techniques such as radio-frequency (RF) data analysis and texture analysis able to quantitatively characterise the radiology data. Although RF data analysis is more appealing compared to B-mode image analysis from the theoretical point of view, in practice however, the added computational cost of working large volume of RF data is at odds with the practical considerations that few ultrasound scanners today offer access to the RF signal. Consequently, texture analysis is more widely applied to characterise ultrasound images.

Texture analysis has been applied on various types of tissues and organ such as carotid artery, breast, heart, prostate, pancreas, and liver. It has also been applied across a various types of imaging modalities such as ultrasound, MRI, CT, conventional X-Ray, and mammogram.

Despite a large study that has been carried out in the application of texture analysis on medical images, to our knowledge, texture analysis of B-mode images has yet to be applied on the characterisation of ovarian cancer. Hence, the aim of this study was to investigate the feasibility of applying texture analysis to characterise ovarian lesions based on B-mode images.

Chapter 3
Texture Analysis and Phantom Image Acquisition

The aim of this chapter is to describe the key topics related to this study. This includes B-mode ultrasound principles, endo-cavity transducer, texture analysis concepts, and phantom image acquisition processes. Texture analysis software (MaZda 4.6, Institute of Electronics, Technical University of Lodz, Poland), image processing software (Adobe Photoshop 6.0, Adobe System Incorporated), and statistical tests used in this study are also briefly discussed. Specific details of the methodologies and statistical analysis are given in the respective chapters.

Section 3.1 gives an overview of the equipment, software, and statistical tests used in this study. It also describes the main processes involved in texture analysis. In Section 3.2, the physical principles related to B-mode image formation are briefly outlined while Section 3.3 briefly introduces the endo-cavity transducer. In Section 3.4, the texture analysis concepts are briefly described which includes texture analysis approaches and methods, with emphasis given to the methods employed in this study. Section 3.5 describes the ultrasound scanner, the transducers and the test object used for phantom studies. Section 3.6 briefly describes the texture analysis software while Section 3.7 briefly describes the image processing software, used in this study. Section 3.8 outlines the statistical tests employed. Finally, Section 3.9 summarises the key topics discussed in this chapter.

Throughout this chapter, the term '*texture analysis*' refers to the '*texture analysis of the digital image*'.

3.1 Background

In this study, a two-dimensional (2D) transvaginal (TSV) transducer was used to acquire B-mode images for phantom studies, using a Toshiba Xario ultrasound scanner. The clinical images used for the classification of normal tissue, cyst and malignant tissue, which were obtained from the achieved images, also were acquired using 2D TSV transducer.

As discussed in Chapter 2, texture analysis techniques, which offer objective characterisation of the images and has gained great interest among researchers, was employed in this study. In general, texture analysis can be broken down into three major steps:

(i) Image acquisition
   - The acquisition of the image by means of a suitable scanner and stored in digital format.

(ii) Image analysis
   - The computation of a characteristic descriptor of numerical parameters for a defined region of interest (ROI), which is able to numerically describe its texture properties.

(iii) Texture / image classification
   - The determination, to which classes (such as normal and abnormal tissue) a texture of defined ROI belongs to.

Figure 3.1 illustrates, in block-diagram form, the texture analysis pathways employed in this study which includes information about the equipment and software used. Prior to features extraction, the original image was first enhanced to improve the visualisation using Adobe Photoshop 6.0. *Magnetic lasso tool* was used to aid in the ROI delineation, and MaZda 4.6 (dedicated software for texture analysis) was used for features extraction.

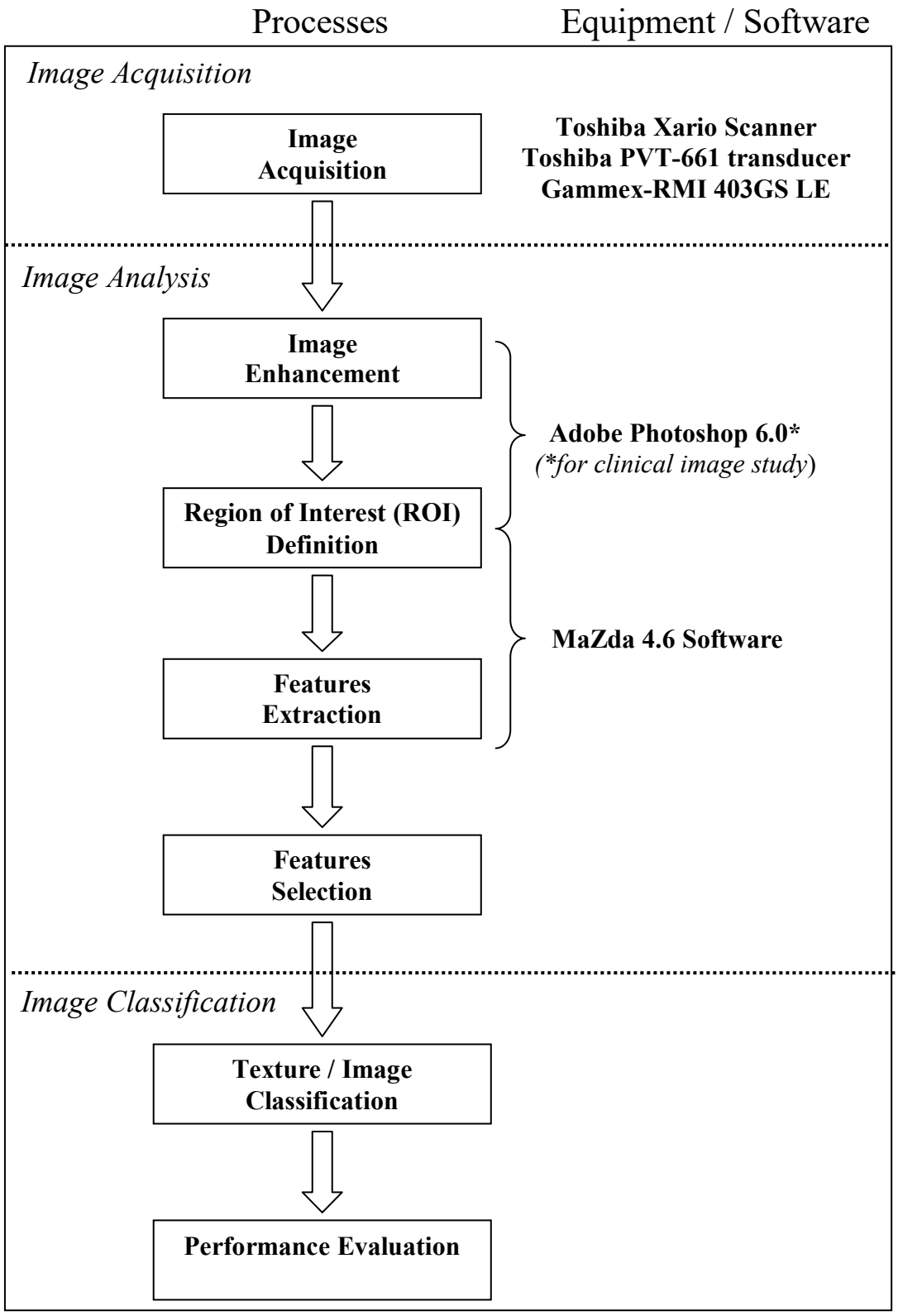


Figure 3.1: Block diagram of the texture analysis pathways employed in the study.

3.2 Basic Principles of B-mode Ultrasound

The use of ultrasound as a diagnostic modality came after Langevin made use of a pulse-echo technique called SONAR (Sound Navigation and Ranging) in 1916. Some of the major advantages of ultrasound, in comparison to other imaging modalities, are due to it being non-invasive, readily available and inexpensive. Real-time B-mode scanning, a most widely used ultrasound technique, did not become widely established until the late 1970s (Whittingham, 2000). According to Barnett (2000), simple B-mode imaging is not capable of producing harmful temperature increases in tissue.

Diagnostic ultrasound is performed by using a pulse-echo approach. A small, spatially localized pulse of ultrasound is produced by a device called a *transducer* and is transmitted into the adjacent tissue. All transducer elements have the same basics components: a piezoelectric plate, a matching layer and a backing layer, as shown in Figure 3.2.

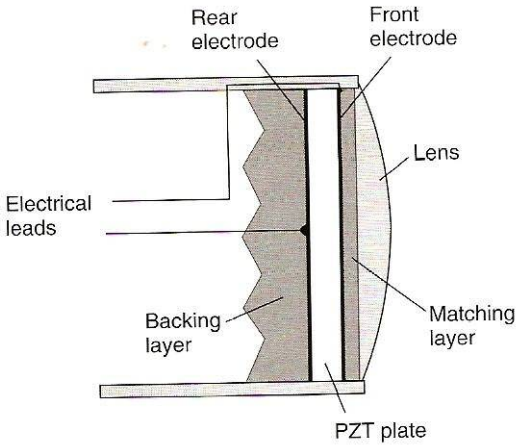


Figure 3.2: The basic component elements in an ultrasound transducer *(adopted from Whittingham, 2007)*

When a voltage is applied to the crystal, it causes it to vibrate and produces the ultrasound pulse, i.e. the transducer converts electrical energy into the transmitted sound waves. In medical diagnostic applications, the frequency range used are between 1-20 MHz (Duck, 2000), which is greater than the upper limit of human hearing. Nevertheless, frequencies up to 40 MHz may be used in special applications and in research (Martin and Ramnarine, 2010).

The ultrasound pulse traverses a straight path and is often referred to as an *ultrasound beam*. The speed of propagation of a sound wave is determined by the medium it is travelling. The properties of the medium which determine the speed of sound are density and stiffness (Martin and Ramnarine, 2010). Density, $\boldsymbol{\rho}$ (rho), is a measure of the mass, for a given volume. It is measured in units of kg m^{-3}. Stiffness is a measure of how well a material resists being deformed when it is squeezed. Stiffness, $\boldsymbol{k}$, is measured in units of Pascal (Pa). Mathematically, the speed of sound is expressed by the following equation:

$$\text{Speed of sound } (\boldsymbol{c}) = \text{Sqrt}(\boldsymbol{k}/\boldsymbol{\rho}) \qquad \text{Equation (2)}$$

Hence, low density and high stiffness lead to high speed of sound. On the other hand, high density and low stiffness lead to low speed of sound. It should be noted, although gasses have low density (which means higher speed is expected in gasses), they have very low stiffness (high compressibility), leading to relatively low speed of sound compared to liquids and solids. Ultrasound pulses travel through biological tissues with an average velocity of about 1,540 m/s. Examples of the speed of sound in some human tissues are given in Table 3.1.

Table 3.1: Speed of sound in some human tissues and liquids (*adopted from Duck, 1990*).

Tissue / organ	Speed (m s^{-1})
Liver	1,578
Kidney	1,560
Amniotic fluid	1,534
Fat	1,430
Water	1,480
Bone	3,190-3,406
Air	333

When an ultrasound wave propagates through soft tissue, the energy associated with the wave is gradually lost – an effect known as attenuation (Martin and Ramnarine, 2010). An ultrasound wave can be attenuated by several mechanisms: absorption, scattering, and divergence of the beam. The rate at which the intensity is attenuated is referred to as the attenuation coefficient. It is measured in units of $dBcm^{-1}$. Examples of the attenuation of sound in some human tissues and liquid are given in Table 3.2.

Table 3.2: Attenuation of sound in some human tissues and liquid (*adopted from Martin and Ramnarine, 2010*).

Tissue / organ	Attenuation (dB cm^{-1} MHz^{-1})
Liver	0.399
Brain	0.435
Muscle	0.57
Blood	0.15
Water	0.02*
Bone	22.00

* proportional to f^2

Another important concept in ultrasound image formation is the *acoustic impedance*. The acoustic impedance of a medium **(*z*)** is a measure of the response of the particles of the medium in terms of their velocity **(*v*)** to a wave of a given pressure **(*p*)** (Martin and Ramnarine, 2010). It is analogous to electrical impedance (or resistance, ***R***). Mathematically, the acoustic impedance is expressed by the following equation:

$$\text{Acoustic impedance } (z) = p/v \qquad \text{Equation (3)}$$

Similar to the speed of sound, the acoustic impedance of a medium is determined by its density **(*ρ*)** and stiffness **(*k*).** Mathematically, it is expressed by the following equation:

$$\text{Acoustic impedance } (z) = \text{Sqrt}(\rho k) \qquad \text{Equation (4)}$$

By combining equation 2 and equation 4, acoustic impedance (*z*) can be rewrite as:

$$\text{Acoustic impedance, } z, = \rho c \qquad \text{Equation (5)}$$

When a sound wave travelling through one medium meets an interface with a second medium of different acoustic impedance, some of the wave is transmitted into the second medium, some is reflected back, and some is scattered. Ultrasound waves are reflected at large interface and scattered by small target.

As the pulse travels deeper into the body, there will be a long train of echoes en route back toward the transducer. In B-mode imaging, the proportion of the wave reflected (echoed) by various tissue interfaces is received by the transducer, converted into electrical energy, amplified and finally displayed as a grey-scale image on a video screen. Pulses for successive beam lines will be transmitted after all of the echoes from the previous beam line have been detected by the transducer. After all of the echoes from the entire beam lines have been detected and processed, these signals are mapped

to the proper locations in the image pixel matrix, and the complete B-mode image is displayed.

An ultrasound image is composed of an array of pixels of different intensity. Each pixel represents a discreet intensity and it relates to the reflection of the ultrasound beams. The different reflectivity of various structures encountered by the pulse cause a corresponding variation of the detected echo strength. Thousands of pixels collectively form the image. It is represented by one of 256 (0 to 255) shades of grey ranging from black (represented as '0') to white (represented as '255'). Figure 3.3 depicts the basic component / process involved in ultrasound imaging.

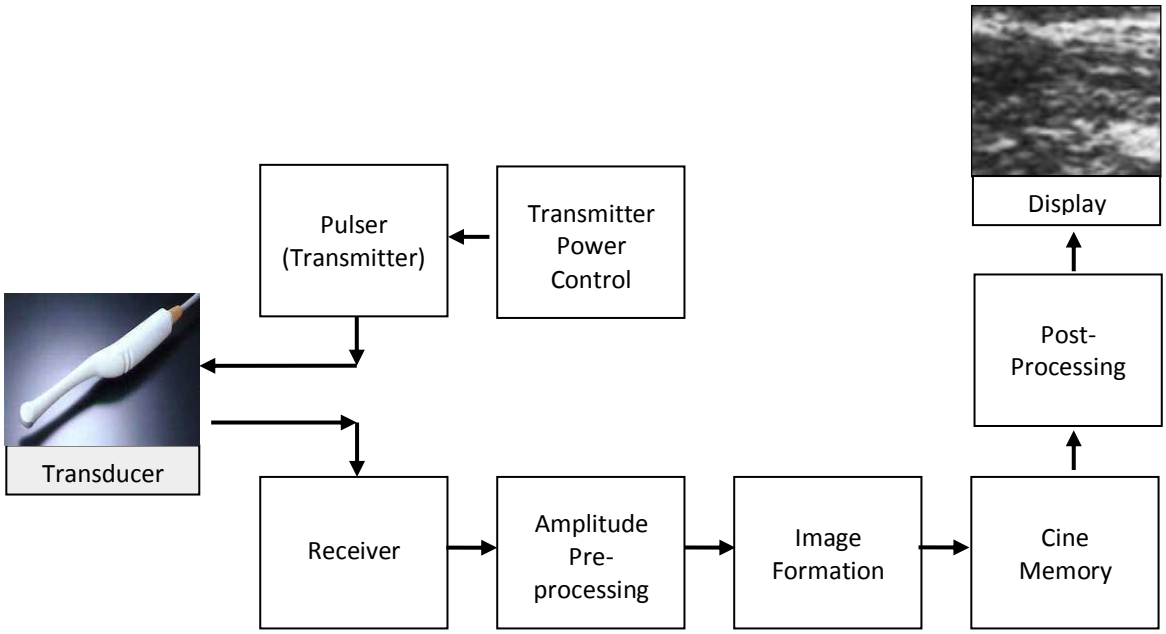


Figure 3.3: Block diagram of a typical B-mode imaging system (*adopted from Martin, 2010*).

In B-mode images, more reflective structures appear brighter than less reflective structures. Liquid structures (e.g., ovarian follicles) do not reflect sound waves and appear black on the viewing screen. Dense tissues (e.g., bone) reflect most of the waves and appear white. Other tissues (e.g., ovarian stroma) are seen in shades of grey depending on its acoustic impedance. Figure 3.4 shows example of B-mode ovarian image obtained using transvaginal (TSV) transducer.

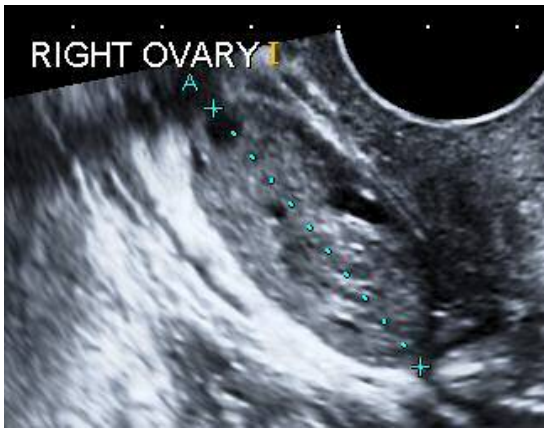


Figure 3.4: Example of B-mode image of the ovary.

3.3 Endo-cavity Transducers

There are several types of endo-cavity transducer: curvilinear, 'end-fire' curvilinear-array, bi-plane trans-rectal, and trans-oesophageal. They are intended for insertion into a natural body cavity or through a surgical opening. Endo-cavity transducer allows the transducer to be placed close to the target organ or mass. Examples of endo-cavity transducers are given in figures 3.5(a) to 3.5(d). In this study, a curvilinear transvaginal transducer (Toshiba PVT-661) was used for phantom study (see section 3.5.2). The Royal College of Obstetricians and Gynaecologists has suggested that the ovarian cysts should normally be assessed using transvaginal ultrasound, as this appears to provide better quality (more detail) than the transabdominal method (RCOG, 2003).

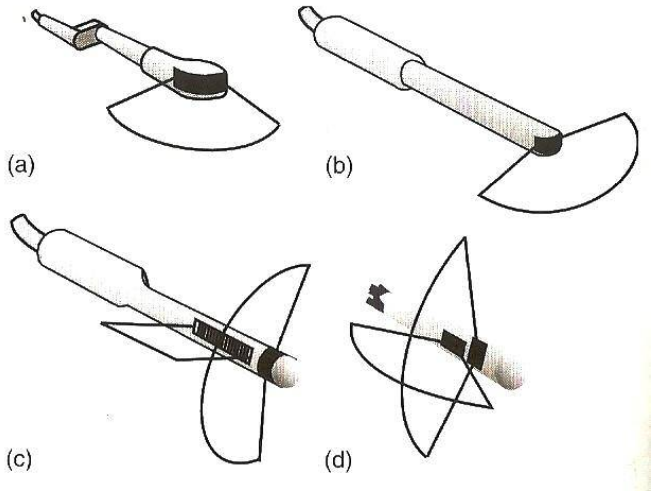


Figure 3.5(a)-3.5(d): Examples of endo-cavity transducers: (a) Curvilinear transducer for trans-vaginal scanning; (b) 'End-fire' curvilinear-array transducer for trans-rectal and trans-vaginal scanning; (c) Bi-plane trans-rectal transducer - allowing both transverse and longitudinal scans of the prostate; (d) Trans-oesophageal transducer with two phased arrays set at right angles (*adopted from Whittingham and Martin, 2010*).

The quality of ultrasound image is correlated to its capability to represent the real anatomy of the structures being examined as closely as possible (Lagalla and Midiri, 1998). This capability among others depends on the axial spatial resolution, the lateral spatial resolution, and the contrast resolution.

The axial spatial resolution refers to its capability to represent separately two objects which are arranged in series along the axis of the beam. This parameter is closely dependent on the transducer frequency: the higher the frequency, the better the axial resolution. An increase in frequency is, however, traded off by a decrease in the depth of beam penetration.

The lateral spatial resolution refers to its capability to discriminate between two objects situated at the same depth on a line perpendicular to the ultrasound beam. This parameter is closely dependent on the dimensions of the ultrasound beam. It can be improved using a narrow beam being at an optimum in the focal zone (Lagalla and Midiri, 1998).

The contrast resolution refers to its capability to distinguish echogenicity differences between neighbouring tissue regions. This parameter is closely dependent on the spatial resolution and the image noise.

Like other types of endo-cavity transducer, the advantage of the transvaginal transducer is that it can be placed close to the target organ or mass. This means, there is less attenuation from intervening tissue, which in turn, a higher frequency may be used, providing a superior axial resolution. The image distortions, and artefacts due to any

tissue heterogeneity or strongly reflecting/refracting interfaces between the transducer and the target, are also reduced (Whittingam and Martin, 2010). Transducer miniaturization continues to increase the capabilities of intra-cavitary, intra-urethral or intra-vascular ultrasound investigation (Bridal *et al*., 2003).

3.4 Texture Analysis

In this project, texture analysis technique was employed to objectively quantify the ovarian images. Hence, a brief review of texture analysis concept is given in this section. More details on texture analysis concept can be found in the articles by Haralick *et al*. (1973), Haralick (1979), Tuceryan and Jain (1998), Materka and Strzelecki (1998), Kociołek *et al*. (2001), Castellano *et al*. (2004), and Srinivasan and Shobha (2008).

3.4.1 What Is Image Texture

In general term, texture is regarded as a property that represents the surface and the structure of an object (Srinivasan and Shobha, 2008). It can be divided into two categories: tactile and visual. A tactile texture refers to the immediate tangible feel of a surface, and is associated with the properties such as smooth, rough, and coarse. Visual texture, which contains variations of intensities (Partio *et al*., 2002), on the other hand, are characterised by the tonal (intensities of image pixels) and the structural (spatial distribution of pixel intensities) properties (Michail *et al*., 2007). Visual texture is an important characteristic for the analysis of many types of images (Haralick, 1979; Rangayyan, 2005). According to Kurani *et al*. (2004), texture is one of the most common features used to analyse and interpret medical images. Tactile texture is not of

interest in this study. Hence, the overview given below focuses only on the visual texture.

According to Haralick (1979) and Bharati *et al*. (2004), however, there is no formal or complete definition of texture. It has been described using various definitions by many researchers. For instance, Bharati *et al*. (2004) quoted the definition of texture given by Russ (1999), which defines texture as a descriptor of local brightness variation from pixel to pixel in a small neighbourhood through an image. In IEEE Standard 610.4-1990, texture in image processing is defined as an attribute representing the spatial arrangement of the grey levels of the pixels in a region. Put simply, in the context of this thesis, texture is defined as the spatial distribution of pixel grey value (intensity) of B-mode images.

3.4.2 What Is Texture Analysis

Texture analysis is the term used for methods developed to quantify image texture (Mathias *et al.,* 1999). It is a branch of imaging science that is concerned with the description of image properties by textural features (Nailon, 2010). The purpose of performing texture analysis is to define a set of texture features that will identify the relevant properties of a texture for a defined ROI (Smutek *et al*., 2003). Texture analysis of medical images is an ongoing field of research, with applications ranging from the segmentation of specific anatomical structures, the detection of lesions, to the differentiation between pathological and healthy tissue (Castellano *et al*. 2004).

There exists a variety of different ways (methods) to compute texture features. These methods can be classified according to the approach used to evaluate the inter-relationships of the pixels.

3.4.3 Texture Analysis Approach

Many textural features approach classification have been given by various authors. Nevertheless, the most comprehensive classification of texture analysis approach were given by Materka and Strzelecki (1998), Castellano *et al*. (2004), Bharati *et al*. (2004), and Materka (2006(a)). According to them, based on the method employed to evaluate the inter-relationships of the pixels, texture analysis can be summarised under the following four approaches: statistical-, model-, structural-, and transform- based.

3.4.3.1 Statistical-based

The use of statistical-based texture features is one of the early methods proposed in the machine vision literature (Tuceryan and Jain, 1998) and is considered to be the most widely used method in medical image analysis (Holli et. al., 2010). According to Chan (1998), this approach has been commonly applied in the characterisation of ultrasound images.

Statistical approach can be further sub-divided into three: first order description statistics, second-order description statistics, and higher-order description statistics.

(a) *First-order description statistics*: First-order description statistics deals with the intensity of the pixels and measure the likelihood of observing a grey value at a randomly chosen location in the image. In other words, the occurrence probability of intensity in an image (the image histogram) is the property used

to calculate the texture features (Nailon *et al*., 1996). It does not take into account the relationship between pixels. These depend only on individual pixel values and not on the interaction or co-occurrence of neighbouring pixel values. Example of the first-order description statistics features is the *average intensity* in an image.

The main advantage of this approach is its simplicity through the use of standard descriptors (e.g. *mean* and *variance*) to characterise the texture. Because it does not consider the spatial relationship and the interaction between pixels, however, the power of this method for discriminating between unique textures in certain applications is limited (Nailon, 2010). Huynen *et al*. (1994) reported that first-order description statistics did not seem to give any additional information for the classification of prostate images. In addition, it is also depth dependent due to attenuation of the beam (Alqahtani, 2010). Hence, first-order description statistics was not employed in this study.

(b) *Second-order description statistics*: Unlike first-order description statistics which deals only with the intensity of the pixels, second-order description statistics takes into account both the intensity of the pixels and the interaction with the neighbouring pixels. Second-order description statistics is defined as the likelihood of observing a pair of grey values occurring at the two endpoints of a dipole of random length placed in the image at a random location and orientation (Tuceryan and Jain, 1998). These are properties of pairs of pixel values. In other words, in second-order statistical texture analysis, information

on texture is based on the probability of finding a pair of grey-levels at random distances and orientations over an entire image (Nailon, 2010).

Three second-order description statistics methods are employed in this thesis: grey level co-occurrence matrix (GLCM), run length matrix (RLM), and gradient. These features are briefly explained in section 3.4.4.1, section 3.4.4.2, and section 3.4.4.3 respectively.

(c) *Higher-order description statistics*: Higher-order description statistics deals with the relationship between three or more pixels. Higher-order description statistics was not employed in this thesis.

3.4.3.2 Model-based

Model-based approach generates an empirical model of each pixel in the image, based on a weighted average of the pixel intensities in its neighbourhood. These estimated parameters of the image models are used as textural feature descriptors. One of the disadvantages of this approach is the computational complexity involved in the estimation of its parameters (Castellano *et al*., 2004). However, with modern computer technology, these parameters are now much more easily computed.

Examples of model-based texture analysis approach are autoregressive (AR) models, Markov random fields (MRF), and fractal models. AR features are employed in this thesis and are briefly explained in section 3.4.4.4.

3.4.3.3 Transform-based

Texture properties of the image may be analysed in a different domain. Transform-based (also known as signal processing) methods apply texture filtering for extracting features either in spatial domain or in frequency domain (Li and Meng, 2009). It converts the image into a new form using the spatial and frequency properties of the pixel intensity variations (Bharati *et al*., 2004).

Examples of model-based texture analysis approach are Fourier, Gabor and wavelet transforms. Wavelet features are employed in this thesis. Wavelet features are briefly explained in section 3.4.4.5.

3.4.3.4 Structural-based

The structural approach assumes that textures are composed of well-defined texture primitives (texture elements) ((Tuceryan and Jain, 1998); (Castellano *et al*., 2004)). The properties and placement rules of these texture primitives define the image texture (Bharati *et al*., 2004). For example, a repetitive arrangement of square and triangular shapes can produce a specific texture (Dhawan, 2003).

The advantage of the structural approach is that it provides a good symbolic description of the image; however, this approach is more useful for synthesis than analysis tasks (Castellano *et al*. 2004; Materka, 2006(a)). The disadvantage of this approach is that it is limited in power unless one is dealing with very regular texture (Tuceryan and Jain, 1998). Because of that, structural approach methods appear to be limited in practicality (Gonzalez cited by Bharati *et al*., 2004).

According to Castellano *et al*. (2004), they did not find any example of structural approach method applied to medical image analysis. Hence, structural approach was not employed in this study.

3.4.4 Texture Analysis Methods

This section discusses the methods employed in this thesis: GLCM, RLM, Gradient, AR-model, and wavelet methods.

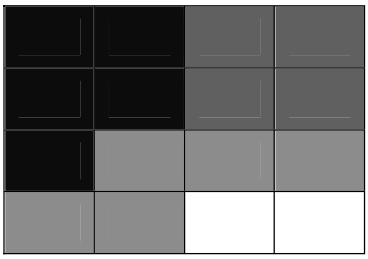
3.4.4.1 Grey Level Co-occurrence Matrix (GLCM)

The GLCM method was first introduced by Haralick for the classification of aerial photographs and satellite images (Haralick *et al*. 1973). It is one of the earliest texture analysers. GLCM has become one of the most well-known and widely used texture features in texture classification (Tuceryan and Jain, 1998).

The GLCM features are computed from intensities of pairs (spatial relationship) of pixels. The spatial relationship is defined in terms of distance (d) and angle (θ) (Sharma and Singh, 2001). It is computed by defining pairs of pixels (i,j) separated by this distance (d), computed across the defined direction (θ). A count is then made of the number of pairs of pixels that possess a given distribution of grey level values. Each entry of the matrix thus corresponds to one such grey-level distribution.

Four matrices can be generated by examining all possible image pixel pairs separated at a certain inter-sample distance (d) in four different directions (θ): 0^0, 45^0, 90^0, and 135^0. MaZda software calculates 11 of 14 textural features as proposed by Haralick at 5

different distance (d = 1-, 2-, 3-, 4- and 5-pixel) giving a total of 220 features. Figure 3.6(a), 3.6(b), 3.7(a), and 3.7(b) illustrates the derivation of GLCM matrices.



(a)

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

(b)

Figure 3.6: Example demonstrating the formation of a co-occurrence matrix from an image. (a) A 4X4 image with four unique grey-level values (ranging from 0 to 3); and (b) Numerical representation of the image (*adopted from Coleman, 2007*).

Grey-level, *i*

Grey-level, *j* / 0^0	**0**	**1**	**2**	**3**
0	4	2	1	0
1	2	4	0	0
2	1	0	6	1
3	0	0	1	2

Co-occurrence matrix with a one-pixel distance in the horizontal direction

(a)

Grey-level, *i*

Grey-level, *j* / 45^0	**0**	**1**	**2**	**3**
0	6	0	2	0
1	0	4	2	0
2	2	2	2	2
3	0	0	2	0

Co-occurrence matrix with a one-pixel distance in the 45^0 direction

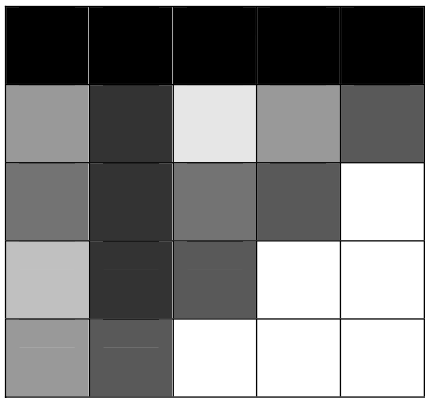
(b)

Figure 3.7: An example demonstrating the formation of a co-occurrence matrix from the image. (a) GLCM in the horizontal direction; and (b) GLCM in the 45^0 direction (*adopted from Coleman, 2007*).

3.4.4.2 ***Run-Length Matrix***

The run-length matrix is a way of searching the image, across a given direction, for runs of pixels (length) that have the same grey-level value (Castellano *et al*., 2004). In simpler words, it is a set of linearly adjacent pixels having the same grey level value (Conners and Harlow, 1980). Grey level runs can be characterized by the grey tone of the run, the length of the run (r), and the direction of the run (θ) (Haralick, 1979).

The run-length matrix describes how many times there are 'j' consecutive pixels with the same grey-level value in a given direction (θ). The length of the run (j) is the number of picture points in the run, and ranges from 2 to n, where n depends on the image size and the direction of measurement. For example, n=5 in horizontal direction for an image with 5X5 pixels. The element $r'(i,j|\theta)$ of the grey level run length matrix specifies the estimated number of times an image contains a run of length (j), for a grey level (i), in the angle θ direction. Figure 3.8(a), 3.8(b), 3.9(a), and 3.9(b) illustrates the derivation of RLM matrices.



(a)

0	0	0	0	0
4	1	6	4	2
3	1	3	2	7
5	1	2	7	7
4	2	7	7	7

(b)

Fig. 3.8: An example demonstrating the formation of a run length matrix from an image. (a) A 5X5 image with eight unique grey-level values (ranging from 0 to 7); and (b) Numerical representation of the image (*adopted from Castellano et al., 2004*).

Grey level (i)	Run length (j)			
0^0	**2**	**3**	**4**	**5**
0	4	3	2	1
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	3	1	0	0

(a)

Grey level (i)	Run length (j)			
45^0	**2**	**3**	**4**	**5**
0	0	0	0	0
1	0	0	0	0
2	3	2	1	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	3	1	0	0

(b)

Figure 3.9: An example demonstrating the formation of a RLM matrix from the image. (a) RLM in the horizontal direction, and (b) RLM in the 45^0 direction (*adopted from Castellano et al., 2004*).

In practice, 4 matrices are computed (horizontal, vertical, and two diagonal directions) (Das *et al*., 2008) and MaZda software computes run-length matrices for these four directions (Szczypinski *et al*., 2009). For each of the matrix, five run-length matrix-based features will be computed giving a total of 20 RLM features. The RLM texture analysis approach characterises coarse textures as having many pixels in a constant grey level run; whereas fine textures is characterised as having few pixels in such a run (Galloway, 1975).

3.4.4.3 Gradient

Another statistical description method is based on the gradient magnitude map of the image. The gradient of an image measures the spatial variation of grey-level values across the image (Castellano *et al*., 2004; Das *et al*., 2008).

For the gradient feature calculation, the following neighbourhood for image pixel *x(i,j)* is defined (Materka, 2006(a)):

G	H	I
L	*x(i,j)*	N
Q	R	S

Figure 3.10: Illustration for gradient feature calculation.
(adopted from Materka, 2006(a)).

The absolute gradient value (ABSV(*i,j*)) is calculated for each 3X3 pixel neighbourhood (figure 3.10):

$$(\text{ABSV}(i,j)) = \text{Sqrt}((R-H)^2 + (N-L)^2) \qquad \text{Equation (6)}$$

Based on the histogram of the image gradient, a similar feature set to the histogram features, the gradient features are calculated for the image intensity distribution: absolute gradient mean, absolute gradient variance, absolute gradient skewness, absolute gradient kurtosis, and percentage of pixels with non-zero gradient.

If, at a point in the image, the grey level varies abruptly from black to white, the gradient texture analysis approach characterises it as having a high-gradient value at that point; whereas, if it varies smoothly from a dark grey to a slightly lighter grey it is

characterised as having a low-gradient value at that point (Castellano *et al*., 2004; Das *et al*., 2008). The gradient may be positive or negative, depending on whether the grey level varies from dark to light or from light to dark. In general, however, because what is of interest is whether we have an abrupt or a smooth grey level variation, the absolute gradient is used (the sign is not taken into consideration) (Castellano *et al*., 2004).

3.4.4.4 Auto-Regressive (AR) Model

The auto-regressive (AR) model uses linear estimates of grey level in texture pixels for texture feature description. It assumes a local interaction between an image pixel and its neighbouring pixels. The assumption made is that every pixel in the image has its grey-level value characterised by the grey-level values of the surrounding pixels: a weighted sum of the grey-level values of the surrounding pixels (Materka, 2006(a)). In other words, it is a way of describing shapes within the image, by finding relations between groups of neighbouring pixels (Castellano *et al*., 2004).

If the grey-level value of a two-dimensional image varies fast, that is, if there are many variations within a small piece of the image, then we associate a high spatial frequency to this part of the image (Castellano *et al*., 2004). On the other hand, if the grey level value varies slowly, then the region has a low spatial frequency.

3.4.4.5 Wavelet

The transform method of texture analysis implemented in MaZda is based on wavelet transform. Wavelet transform is a tool that separates data into different frequency components (Materka, 2006(b)). Wavelet transform is computed with a cascade of filtering: high- ,H, and low- ,L, pass filters. For two-dimensional images, the transform is performed firstly for all image rows, and then for all image columns (Kocoilek *et al*., 2001). This yields four different sub-bands: L-L, L-H, H-L, and H-H (figure 3.11).

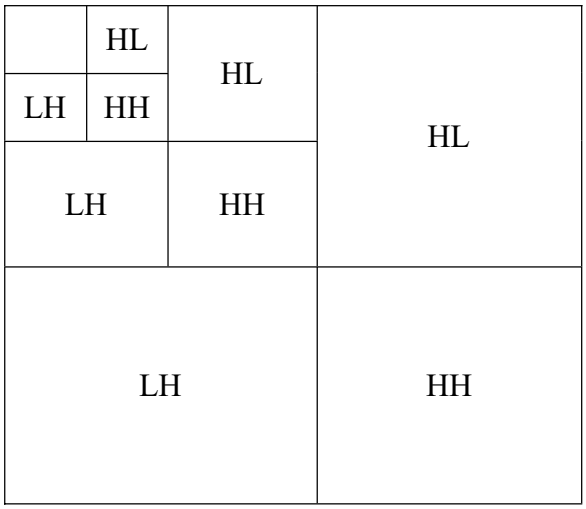


Figure 3.11: Sub-band image
(*adopted from Kociolek et al., 2001*).

For a given image, MaZda calculates up to a maximum of 8 scales. It is calculated only if output sub-bands have dimensions at least 8 by 8 points. This results in a set of wavelet coefficients that correspond to different scales.

The Wavelet transform has been used for image segmentation and classification (Bharati *et al*., 2004), and widely found in the analysis of medical images (Livens *et al*.,

1997). According to Bharati *et al*. (2004), this method has been preferred recently in image texture analysis due to its space-frequency decomposition abilities. Wavelet feature was employed in this study.

3.5 Phantom Image Acquisition

3.5.1 Ultrasound Scanner

It was not until the late 1960s that the first commercial ultrasound scanner became available for clinical use, and its role as a diagnostic imaging tool is increasingly important (Board of the Faculty of Clinical Radiology, 2005). In this study, a Toshiba Xario (Toshiba Medical System, Europe) (Figure 3.12) was used to acquire the images for phantom study. The specifications of this scanner are given in Appendix A.

Figure 3.12: Toshiba Xario Scanner.

3.5.2 Transvaginal Transducer

As discussed in section 3.3, one of the advantages of transvaginal transducer is that it can be place close to the target organ or mass and is widely used for ovarian cancer diagnosis. In this study, a 6.0 MHz Toshiba transvaginal transducer model PVT-661VT (figure 3.13) was used to acquire the images for phantom study.

Figure 3.13: Toshiba PVT-661VT transducer.

3.5.3 Test Object

The phantom study was performed using the Gammex-RMI model 403GS LE (Gammex-RMI Ltd., Nottingham) general purpose test object (figure 3.14). The tissue-equivalent material used in the phantom is an evaporated-milk-based gel, which is ultrasonically similar to human tissue and allows the use of normal scanner control settings (RMI 403LE GS manual). The use of tissue-equivalent test phantom gives the measurement results that are consistent with clinical performance (Browne *et al.*, 2003). The RMI model 403GS LE phantom provides the measurement up to 16 cm depths. The acoustic properties of this phantom are given in table 3.3.

Figure 3.14: Gammex-RMI model 403GS LE phantom.

Table 3.3: Acoustic data of the RMI 403LE GS test phantom tissue-equivalent material.

Tissue-equivalent material	**Speed of sound (m/s)**	**Attenuation coefficient (dB/cm-MHz)**
Evaporated milk gel Model 403GS LE	1540 @ 2–18 MHz	0.7 @ 2–18 MHz

3.6 MaZda Software

In this study, MaZda 4.6 (Institute of Electronics, Technical University of Lodz, Poland) was used to extract the textural features. Brief explanations for this software are given in this section. More details about this software are available on-line at *http://www.eletel.p.lodz.pl/mazda/*.

3.6.1 Background

MaZda is an acronym derived from 'Macierz Zdarzen' which means co-occurrence matrix in Polish. It was developed in 1998 specifically for the analysis of image textures (Szczypinski *et al*., 2007). The package includes two executable files: (i) MaZda (image processing and computation of textural features), and (ii) B11 (for data visualisation, classification and segmentation). Initially, it was aimed at analysis of MRI images. However, it has also been widely used for other types of images such as x-ray, CT, and ultrasound images.

In general, MaZda loads images in the form of Windows Bitmap files, DICOM files, and unformatted grey-scale image files (raw images) with pixel intensity encoded with 8 or 16 bits. The software has been used for many different purposes such as for MRI measurement protocol optimisation, for texture analysis of medical images, for food quality studies, and for other purposes (Szczypinski *et al*., 2009).

MaZda software has been widely used for texture analysis in medical domain. This involves various types of diseases and imaging modalities. Table 3.4 summarises some of previous studies that employed MaZda software to extract the textural features.

Table 3.4: Example of the medical domain studies that used MaZda software for features extraction.

Author / Year	Area of study
Harrison *et al*. (2010)	Classification of white matter and multiple sclerosis lesions.
Mayerhoefer *et al*. (2010)	Assessment of biochemical changes in meniscal tissue.
Alqahtani *et al*. (2010)	To investigate the effect of varying the gain and the dynamic range setting on the texture features in-vivo using a 3D ultrasound probe
Abella *et al*. (2009)	The quantification of cellular infiltrate in lung tissue images in an allergic asthma.
Zhang *et al*. (2008)	To discriminate between multiple sclerosis (MS) lesions, normal appearing white matter (NAWM) and normal white matter (NWM).
Mescam *et al*. (2007)	To study the relationship between physiological parameters and textural parameters of the liver.
Hajek *et al*. (2006)	To distinguish brain tumours.
Hollingsworth and Lomas (2005)	To discriminate between cirrhotic and healthy livers tissue.
Chen *et al*. (2005)	For the evaluation of anti-vascular therapy of mammary carcinomas in mice.
Lerouxel *et al*. (2004)	For osteoporosis detection based on bone images.
Bonilha *et al*. (2003)	For the detection and quantification of hipocampal sclerosis.
Jirak *et al*. (2002)	Discrimination of healthy and cirrhotic livers.

3.6.2 Image Normalisation

Before the texture features are computed / derived for the co-occurrence matrices, the run length matrices, and the image gradient features, the image fragments inside the ROIs are first normalised. The purpose of performing normalisation is to reduce the variation in the image due to difference in scanner settings (Chan and McCarty, 1990).

MaZda provides three options for image normalisation: (i) default (analysis is made for original image); (ii) ±3σ (image mean value, μ, and standard deviation, σ, is computed, then analysis is performed for grey scale range between μ -3σ and μ +3σ); and (iii) 1%-99% (grey-scale range between 1% and 99% of cumulated image histogram is taken into consideration during analysis).

Collewet *et al*. (2004) have carried out a study to compare the influence of several normalisations of grey level methods on the discriminatory power of texture analysis. They found that '±3σ' method gives the best results, whereby it is able to enhance the differences between the two sample groups tested. This method of normalisation ('±3σ') has been used by Harrison *et al*. (2008) in the texture analysis of non-Hodgkin lymphoma. Hence, the image grey-level normalisation method, which normalises image intensities in the range ($\mu - 3\sigma$, $\mu + 3\sigma$), was used in this study.

3.6.3 Region of Interest (ROI) Selection

To analyse image properties in a selected area and not in the surrounding tissue, the image segment corresponding to the organ must be defined as the ROI for the analysis. Region of interest (ROI) is a set of pixels in 2D images selected to be analysed. The MaZda software allows the ROI to be drawn of arbitrary shape. In this study, in order to improve the ROI delineation process, the ROI was drawn based on the temporary segmented image, which was outlined using *magnetic lasso tool* (see section 3.7).

3.6.4 Features Extraction

Texture features, i.e. the outputs of texture analysis, are mathematical parameters (numerical) computed from the distribution of pixels, which characterise the texture type, and thus, the underlying structure of the objects shown in the image (Castellano *et al*., 2004).

MaZda software computes six types of textural features: Histogram-based (first-order statistics), GLCM, RLM, Gradient, AR, and Wavelet. Histogram-based features however, were not included in this study (see section 3.4.3.1(a)). Findings from Bader *et al*. (2000) showed that grey level histogram-based texture features extracted from ultrasound images were not able to distinguish benign tumours from carcinomas. Similarly, in the study on thyroid gland, Catherine *et al*. (2006) reported that first-order description statistics did not prove to be representative of the image characteristics. Moreover, first-order description statistics also found to be depth dependent (Alqahtani, 2010).

3.7 Adobe Photoshop Software

In this study, Adobe Photoshop 6.0 (Adobe System Incorporated) was used to enhance the image, particularly to aid in the ROI delineation for the clinical images. Brief explanations of this software are given in this section.

In scientific research, images fall broadly into two categories: the original image and the corrected image. The corrected image is often referred to as 'enhanced'. Post-processing is required because of limitation in imaging devices and associated instrumentation (Sedgewick, 2008). This limitation includes, among others, image blurring and image noise. Post-processing, which results in enhanced images, is often used to improve the visualisation of the details in the image. The purpose of performing image enhancement using Adobe Photoshop 6.0 in this study was to produce an image, in which the area of interest is adequately visualised and separated from the background during ROI delineation step.

Besides being used to improve the visualisation, Adobe Photoshop 6.0 also was employed in this study to aid in the outlining the ROI (using *magnetic lasso tool*). According to Park *et al*. (2005), the use of *magnetic lasso tool* is appropriate for semiautomatic segmentation of the anatomical structures in the MRIs, CTs, and other medical images, and is expected to be widely used on the personal computer.

The use of this software for image enhancement and ROI delineation is given in more detail in the respective sections in chapter 6 and chapter 7.

3.8 Statistical Analysis

There are several statistical tests used to evaluate the results in this study, depending on the intention of the analysis being carried out. For example, statistical test is used to quantify the agreement between two operators, to measure the reliability of the image texture produced, to compare the mean between the two groups, and to assess the influence of one variable to the other. This section explains in brief those statistical tests. Bland-Altman plot, a graphical approach that plots differences against means, is also briefed.

In this study, statistical analysis was performed using SPSS 16.0 for Windows (SPSS Inc, Chicago, Illinois, USA).

3.8.1 Coefficient of Variation (CoV)

Coefficient of variation (CoV) is the standard deviation (SD) divided by the mean, often multiplied by 100 to give percentage (Bland, 2000). It is used to measure the consistency of the data (the uniformity in the values of the data from the mean). The advantage of CoV is that it is independent of the units of observation (Kirkwood and Sterne, 2003). A distribution with smaller CoV is taken as more consistent than that with larger CoV. The coefficient of variation (CoV) is widely used to measure random error (Dudley, 2010).

In this study, CoV was used to determine the variation (consistency) in the derived texture features: ultrasound system repeatability, intra-operator repeatability, inter-operator reproducibility, clinical image repeatability, influence of ROI size on the

variance of the mean, and ROI delineation repeatability (see chapter 4, chapter 5, and chapter 6).

3.8.2 T-Test and Wilcoxon-Mann-Whitney (WMW) Test

Two-sample t-test is used to compare the mean value between two groups for the normally distributed data while Wilcoxon-Mann-Whitney- test (also called the Mann–Whitney U test or Wilcoxon rank-sum test) is used for the non-normally distributed data (Jones, 2009; Kirkwood and Sterne, 2003). A p-value of <0.05 is considered as significant difference. A Q-Q plot is used to assess the normality of a distribution. For a normal distribution, the plot should show a linear relationship (Leeds Metropolitan University, 2009).

In this study, t-test and WMW test was used to assess the influence of the 'ultrasound calliper line' as well as for comparing the texture features values between different ovarian tissue types (see chapter 6 and chapter 7).

3.8.3 Correlation Coefficient

The goal of a correlation analysis is to see whether two measurement variables co-vary (McDonald, 2009). It is useful to measure the strength of relationship (how one changes with the other) between two variables (Bland, 2000) and provides the direction of the relationship (Jones, 2009). If the slope is significantly different from zero, there is a significant relationship between the two variables: as the values of one variable increase, the values of the other variable either tend to increase (positive correlation) or tend to decrease (negative correlation) (McDonald, 2009).

In this study, correlation test was used to determine whether there is a significant relationship between the derived features and the three possible influence factors: gain setting, ROI size, and ROI depth (see chapter 4).

3.8.4 Receiver Operating Characteristics (ROC) Curve Analysis

ROC curves assess the value of a diagnostic test by providing a standard measure of the ability of a test to correctly classify subjects (Morrison, 2005). It has been used extensively in the biomedical field to assess the efficacy of diagnostic tests in discriminating between healthy and diseased individuals (Morrison *et al*., 2003), and is particularly useful when comparing two or more measures (Altman and Bland, 1994). An area under the curve (AUC) close to 1 indicates a strong discriminatory power / ability of the indicator variable while AUC close to 0.5 indicates that the variable has little discriminatory power (Morrison, 2005). Simon (2008) divides the ability of the discriminator into five categories, based on the area under the ROC curve: excellent, very good, good, fair, and poor for the area value of >0.97, 0.92 to 0.97, 0.75 to 0.92, 0.50 to 0.75, and <0.50 respectively.

In this study, ROC curve analysis was used to determine the discriminatory ability of the tested textural features in discriminating the three groups: normal ovary, cysts, and malignant tissue (see chapter 7).

3.8.5 Bland-Altman Plot

Bland-Altman plot is a scatter plot of variable means, plotted on the horizontal axis and the differences plotted on the vertical axis, which shows the amount of disagreement between the two measures (via the differences). The graphical approach that plots differences against means is the most informative approach (Bassani *et al*., 2007). The presentation of the 95% limits of agreement is for visual judgement of how well two methods of measurement agree: the smaller the range between these two limits the better the agreement is (Myels and Cui, 2007). This plot is also useful for evaluating whether there is any systematic difference between the methods or whether the degree of random variation changes with the mean value (Laugsand *et al*., 2010).

In this study, Bland-Altman plot was used to show the amount of disagreement between the two operators (see chapter 5).

3.9 Summary

Texture analysis technique was employed in this study to provide an objective interpretation of the B-mode image. To achieve that, MaZda 4.6 software was used for features extraction. A Toshiba Xario ultrasound scanner was used to acquire B-mode images for phantom studies, using a transvaginal transducer (Toshiba PVT-661VT). For the clinical image study, Adobe Photoshop 6.0 was used to improve the visualisation of the area of interest by applying *gradient-filter* to the original image. Adobe Photoshop 6.0 was also used to aid in the ROI delineation utilising another feature called *magnetic lasso tool*.

Texture analysis characterises a defined region in an image by its texture content. In general, texture analysis process can be broken down into three major steps namely image acquisition, image analysis, and texture / image classification. Texture analysis is classified according to the approach used to evaluate the inter-relationships of the pixels. The most comprehensive classification of texture analysis summarises it into four groups: statistical-, structural-, model-, and transform- based. These approaches can be further sub-divided into several methods.

Five texture analysis methods were employed in this study. This includes GLCM, RLM, gradient, AR, and wavelet. GLCM, RLM, and gradient belong to the statistical-based approach, while AR and wavelet belong to the model-based and transform-based approach, respectively.

Several statistical tests were employed to evaluate the results in this study. This includes coefficient of variation (CoV), t-test and Mann-Whitney-Wilcoxon (WMW)

test, correlation analysis, and receiver operating characteristics (ROC) analysis. Apart from that, Bland-Altman plot was also used for visual interpretation of the inter-operator reproducibility assessment.

Chapter 4
Influence of ROI Size, ROI Depth, and Gain Settings on Image Texture

The objective of this chapter is to assess factors that potentially affect the extracted texture features, namely the ROI size, the ROI depth and the gain setting. Section 4.1 reviews the influence of those three factors on the extracted features. The methodology and statistical analysis are explained in Section 4.2. Section 4.3 presents the results, and these are discussed in Section 4.4. Finally, Section 4.5 summarises this chapter.

4.1 Background

It was reported that besides the variations in the imaged tissue itself, the extracted texture features may also vary due to other factors. Castellano *et al.* (2004) pointed out that the effect of external factors on some texture parameters must be taken into consideration before using texture analysis techniques. If this is not addressed then, according to Chan and McCarty (1990), the value obtained may not reflect the actual texture of the tissue, which may in turn influence the performance of the texture analysis technique when applied for classification.

In previous work carried out by He *et al.* (2004) on a tissue-equivalent phantom, they reported some texture features exhibit a dependency on the size of the area they were extracted from. This dependency can be divided into two: (i) The *variance* of the extracted features - most features exhibited larger variance when small ROIs were used; and (ii) The *mean* of the extracted features - some features exhibit a monotonically increase as a function of ROI size.

Similarly, the influence of the ROI size has been reported in clinical image studies. Using co-occurrence matrix features to discriminate prostatic tissue from ultrasound images, Basset *et al.* (1993) reported that the results from using a smaller ROI were not as good as results using larger ROI, which shows that the size of ROI will affect the performance of texture analysis technique. According to He *et al.* (2004), since the pathological area may vary, and the size of the tumour is normally larger than the size of the normal area, there is a need to determine the appropriate ROI size before extracting the features for image classification purposes.

Besides the ROI size, another factor that needs to be considered during feature extraction is the depth of the ROI. A study carried out by Morris (1988(a)) revealed that some texture features (such as mean grey level and contrast) are dependent on the ROI depth. Although some studies extracted the texture features from a fixed ROI depth, this approach limits the flexibility of the technique because in real-life scenarios the lesions may appear at various depths.

It should be noted that the texture analysis technique is also sensitive to the acquisition condition (Collewet *et al.*, 2004). According to Chan and McCarty (1990), the value of the extracted features may change drastically under different scanner settings, even if the same tissue is being imaged. Examples of such factors are gain and dynamic range (DR). In this study, however, we only looked at the influence of gain setting on the B-mode image texture. Assessment of the influence of DR was excluded from this study because this setting is normally set and very rarely adjusted when doing an examination. Moreover, a study carried out by Alqahtani et al. (2010) showed that varying the DR

yielded a small variation on all texture features tested (gray-level histogram, GLCM, RLM, gradient, AR, and wavelet) and concluded that changing the dynamic range has no significant effect on texture features.

4.2 Methodology and Statistical Analysis

4.2.1 Influence of ROI Size

The objective of this section is to evaluate the influence of the ROI size on the texture features, particularly on the *variance* and the *mean* of the extracted features.

Ten images were acquired using a Toshiba Xario scanner with a 6.0 MHz transvaginal transducer (model PVT-661VT). An RMI model 403 GS LE was used as a test object. To acquire the image, the following procedure was used: a stand was used to hold the transducer throughout the image acquisition period. The machine settings were kept constant by using manufacturer's pre-defined settings, *Endo-Vaginal-Gynae***.** Finally, the images were transferred to the personal computer (PC) in *.bmp* format for further analysis. Figure 4.1 illustrates the experimental setup and the texture analysis process flow. Texture analysis software, MaZda version 4.6, was used to extract the texture features from the images. Five texture features (GLCM, RLM, gradient, AR, and wavelet) were extracted from ten square ROIs (10x10, 20X20, ..., 100X100 pixels), which were placed in the centre of the image as shown in Figure 4.2.

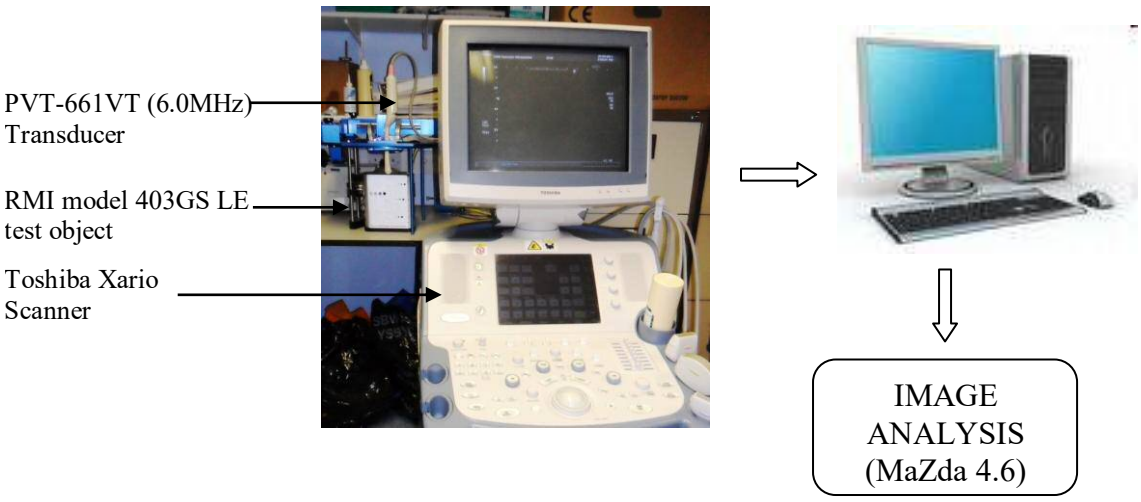


Figure 4.1: Block diagram of ROI size, ROI depth and *gain* setting assessment.

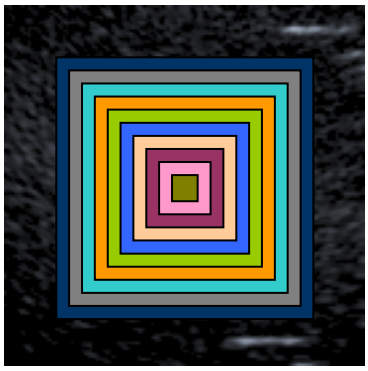


Figure 4.2: Ten square ROIs were placed in the centre of the image for ROI size assessment.

A coefficient of variation (CoV) (see Section 3.8.1) was calculated for each texture feature to determine and compare the variation (consistency) of the features extracted from a particular ROI size. The CoV is a useful statistical measure to quantify the dispersion (variance) of data relative to the mean. Correlation analysis (see Section 3.8.3) was performed to determine the relationship between the ROI size and the *mean* of the extracted texture features, and to see whether there is a significant increase (for the positive correlation case) or decrease (for the negative correlation case) in the *mean*, as the ROI size increases.

4.2.2 Influence of ROI Depth

The objective of this section is to evaluate the influence of the ROI depth on the extracted texture features; i.e. whether as the depth of the ROI increases, the value of the extracted texture features tends to increase/decrease as well.

The procedures carried out for the image acquisition were similar to the ROI size assessment (see Section 4.2.1). The images were acquired for three scanner focus settings (4, 6, and 8 cm). The texture features were then extracted at six different depths (1, 2, 3, 4, 5, and 6 cm) for each focus. An example of ROI placement is shown in Figure 4.3.

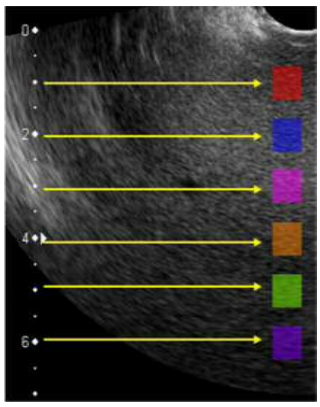


Figure 4.3: Example of ROI placement for depth influence assessment.

Correlation analysis (see Section 3.8.3) was performed to determine the relationship between the ROI depth and the extracted texture features, i.e. whether there is a significant increase (for the positive correlation case) or decrease (for the negative correlation case) as the ROI depth increases.

4.2.3 Influence of Gain Setting

The objective of this section is to evaluate the influence of the scanner *gain* setting on the texture features, i.e. whether as the value of the gain setting increases, the value of the texture features tends to increase/decrease as well.

The images were acquired using a Toshiba Xario scanner with a 6.0 MHz transvaginal transducer (model PVT-661VT). An RMI model 403 GS LE was used as a test object. To acquire the image, the following procedure was used. A stand was used to hold the transducer throughout image acquisition period. All other machine settings (except the *gain*) were kept constant across the image acquisition process. The setting was set to the manufacturer's pre-defined *Endo-vaginal-Gynae* settings. The images were acquired at five different gain settings: 80, 85, 90, 95, and 100 dB. The images were transferred to the PC in *.bmp* format for further analysis. Please refer to Figure 4.1 for the experimental setup and the process flow illustration. A square ROI (50X50 pixels) was drawn in the centre of the image for features extraction using MaZda version 4.6 (figure 4.4).

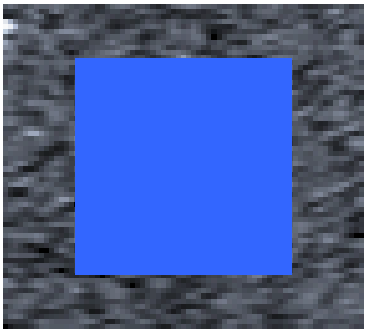


Figure 4.4: A square (50X50 pixels) ROI was placed in the centre of the image for features extraction assessment of the influence of gain setting.

Correlation analysis (see Section 3.8.3) was performed to determine the relationship between the scanner *gain* setting and the extracted texture features, i.e. whether there is a significant increase (for positive correlation case) or decrease (for negative correlation case), as the *gain* increases.

4.3 Results

4.3.1 Influence of ROI size

ROI size may influence on the extracted texture features in two ways: (i) the *variance* of the extracted texture features (within the same ROI size), and (ii) the *mean* of the extracted texture features (between different ROI sizes). The COV, which was used to evaluate the *variance* of the texture features extracted from various ROI sizes are summarised in Table 4.1. The mean and the standard deviation results are presented in Appendix B. Generally, all features exhibit relatively low variations. The results show that GLCM, RLM, AR and wavelet features yield more consistent values, while the gradient feature was most sensitive to the ROI size, particularly for an ROI size of less than 900 pixels. For an ROI size of more than 900 pixels, the extracted values for all features are relatively consistent. The CoVs as a function of ROI size for all the features evaluated are shown in Figure 4.5.

Table 4.1: Summary of the results for the influence of ROI size on the *variance* of the extracted features.

	Coefficient of Variations (%)									
ROI Size (pixels)	**100**	**400**	**900**	**1,600**	**2,500**	**3,600**	**4,900**	**6,400**	**8,100**	**10,000**
GLCM	4.50	3.20	1.72	1.67	1.57	2.14	2.43	1.84	1.53	1.49
RLM	6.21	4.17	2.31	2.31	2.47	3.09	3.04	2.64	2.88	2.57
Gradient	15.51	12.84	8.04	5.16	3.56	4.14	3.15	3.79	3.75	3.69
AR	8.22	2.47	1.86	1.77	0.82	0.63	0.65	0.54	0.51	0.44
Wavelet	8.29	2.55	3.71	0.98	2.59	1.43	1.31	1.44	1.99	0.92

** 1cm = 48 pixels*

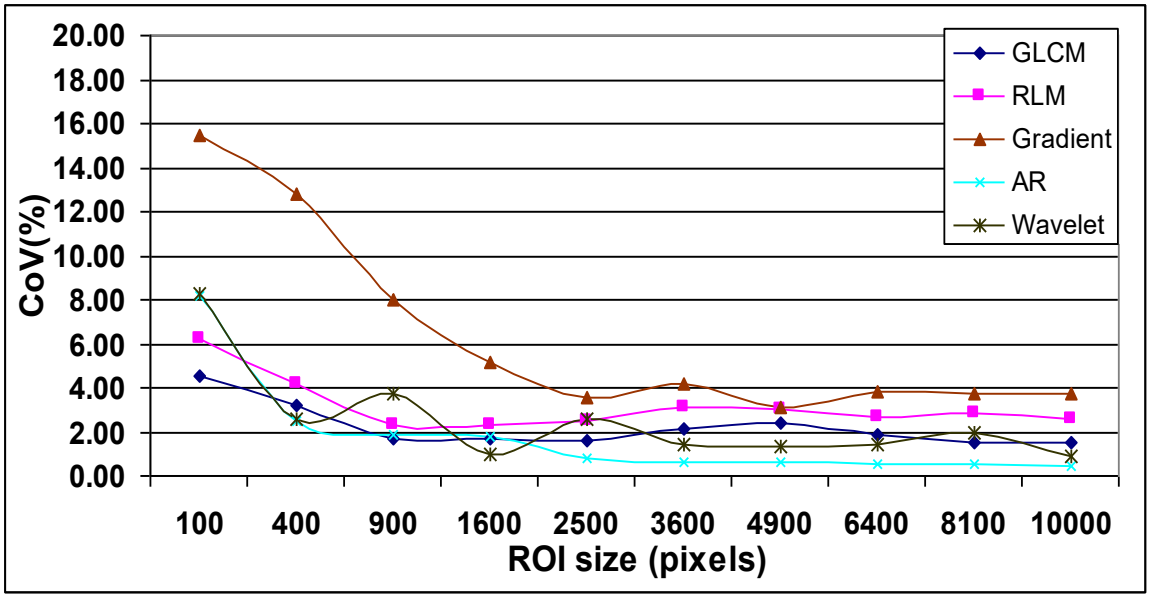


Figure 4.5: The CoVs for ROI size influence on the *variance* of the extracted features.

Figure 4.5 shows that the *variance* of the extracted features for all features exhibits a clear dependency on the ROI size for a size of less than 900 pixels. However, the *variance* is relatively small and consistent when the ROI is greater than 900 pixels.

Figure 4.6 shows an example of the distribution of the texture feature (GLCM) which is not significantly correlated with the ROI size. A similar pattern was also observed for Gradient and Wavelet features. Figure 4.7 shows an example of the texture feature (RLM), which is significantly correlated to the ROI size. The figure clearly shows that RLM feature increases monotonically as the ROI size increases. A similar pattern was also observed for the AR feature. The correlation test results of all five texture features extracted from various ROI size are summarised in Table 4.2, while the mean and standard deviation results are presented in Appendix B. The results given in Table 4.2

show that two of them (RLM, and AR) have a significant correlation with the ROI size, i.e. they vary linearly as the ROI size increases.

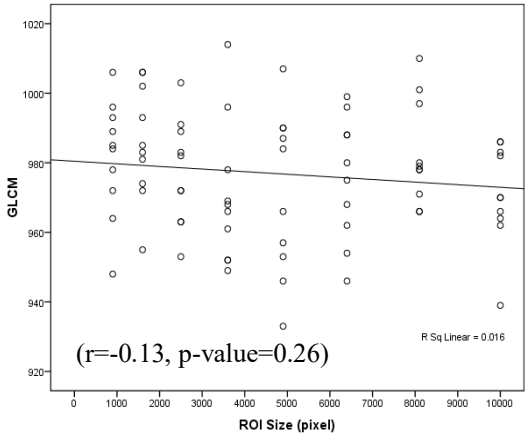


Figure 4.6: Example of scatter plot showing the distribution of the extracted feature (GLCM), which does not has a significant correlation with the ROI size.

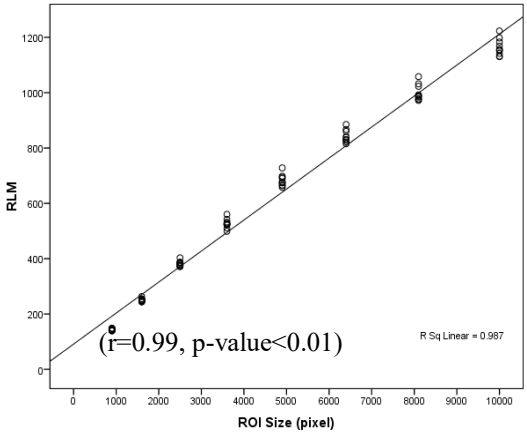


Figure 4.7: Example of scatter plot showing the distribution of the extracted feature (RLM), which has a significant correlation with the ROI size.

Table 4.2: Summary of the correlation test for the influence of the ROI size on the *mean* of the extracted features.

	Correlation coefficient (r)	p-value	Significantly correlated?
GLCM	-0.13	0.26	No
RLM	0.99	<0.01	Yes
Gradient	0.05	0.68	No
AR	0.44	<0.01	Yes
Wavelet	0.20	0.08	No

**ROI size tested between 900 to 10,000 pixels*

4.3.2 Influence of ROI Depth

Figure 4.8 shows an example of the distribution of a texture feature (GLCM) that is not significantly correlated with ROI depth. A similar pattern was also observed for wavelet feature. Figure 4.9 shows an example of a feature (AR) that shows significant correlation between the ROI depth and the extracted feature, showing a clear relationship between the extracted features and the ROI depth. A similar pattern was also observed for RLM and gradient features. Each texture feature exhibits a similar pattern for all three focus settings tested (4, 6, and 8 cm).

The correlation test results of the texture features extracted from various ROI depths are summarised in Table 4.3, while the mean and standard deviation results are presented in Appendix C. The results given in Table 4.3 show that three of them (RLM, Gradient, and AR) have a significant correlation with the ROI depth (vary linearly as the ROI depth increases).

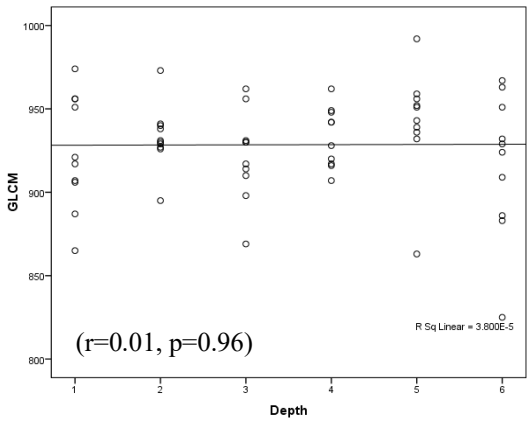


Figure 4.8: Example of scatter plot showing the distribution of the extracted feature (GLCM), which does not have a significant correlation with the ROI depth.

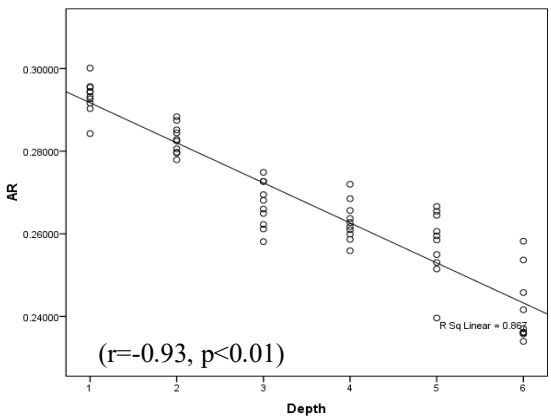


Figure 4.9: Example of scatter plot showing the distribution of the extracted feature (AR), which has a significant correlation with the ROI depth.

Table 4.3: Summary of the correlation test for the influence of the ROI depth on the extracted features.

	Focus = 4 cm		Focus = 6 cm		Focus = 8 cm	
	'r' / p-value	**Sig. corr?**	**'r' / p-value**	**Sig. corr?**	**'r' / p-value**	**Sig. corr?**
GLCM	0.01/0.96	No	-0.14/0.24	No	-0.07/0.57	No
RLM	-0.83/<0.01	Yes	-0.82/<0.01	Yes	-0.81/<0.01	Yes
Gradient	-0.44/<0.01	Yes	-0.58/<0.01	Yes	-.051/<0.01	Yes
AR	-0.93/<0.01	Yes	-0.91/<0.01	Yes	-0.90/<0.01	Yes
Wavelet	-0.17/0.19	No	-0.29/0.12	No	0.11/0.36	No

'r'=correlation coefficient; Sig. corr=Significantly correlated.

4.3.3 Influence of Gain Setting

Figure 4.10 shows an example of the distribution of a texture feature (GLCM) which is not significantly correlated. Figure 4.11 gives an example of a feature (RLM) that shows significant correlation with the *gain* setting and the extracted feature. A similar pattern was also observed for gradient, AR and wavelet features.

The correlation test results of the texture features and various gain settings are summarised in Table 4.4, while the mean and standard deviation results are presented in Appendix D. The results in the Table 4.4 show that for gain settings ranging from 80 dB to 100 dB, only the GLCM feature does not significantly correlate with the gain setting. The other four (RLM, Gradient, AR, and wavelet) have a significant correlation with the gain setting (vary linearly as the gain increases).

However, when the correlation test was confined to changing the gain from 90 dB to 100 dB, four of the evaluated features (GLCM, Gradient, AR, and wavelet) did not significantly correlate with the gain. Only the RLM feature shows a significant correlation with this gain setting.

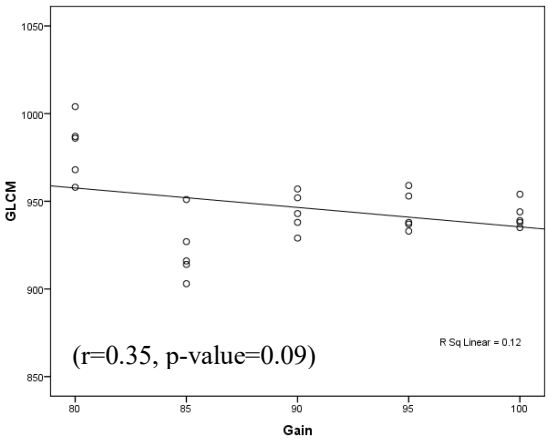


Figure 4.10: Example of scatter plot showing the distribution of an extracted feature (GLCM), which does not have a significant correlation with the gain setting.

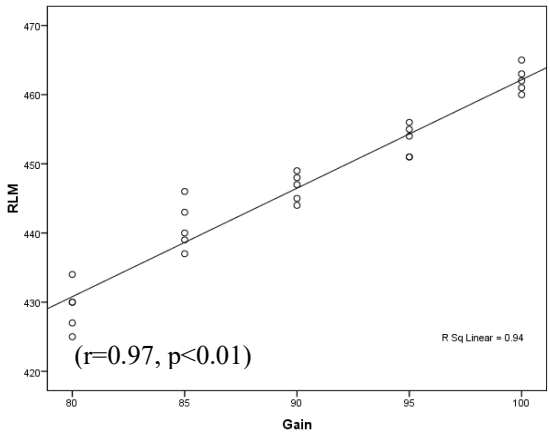


Figure 4.11: Example of scatter plot showing the distribution of an extracted feature (RLM), which has a significant correlation with the gain setting.

Table 4.4: Summary of the correlation test for the influence of the gain on the extracted features.

	Gain (dB): 80,85,90,95,100			**Gain (dB): 90,95,100**		
	'r'	**p-value**	**Sig. corr?**	**'r'**	**p-value**	**Sig. corr?**
GLCM	-.035	0.09	No	-0.08	0.78	No
RLM	0.97	<0.01	Yes	0.96	<0.01	Yes
Gradient	-0.73	<0.01	Yes	0.09	0.76	No
AR	-0.72	<0.01	Yes	-0.31	0.26	No
Wavelet	0.82	<0.01	Yes	0.36	0.18	No

'r'=correlation coefficient; Sig. corr=Significantly correlated.

4.4 Discussion

The success of applying the texture analysis technique is subject, among other factors, to the reliability of the extracted features. Three factors associated with the reliability were considered in this chapter: (i) ROI size, (ii) ROI depth, and (iii) gain setting.

In this study, the smallest ROI investigated was 100 pixels. According to Lerski (2006), most methods of texture analysis cannot produce useful representative data extracted from ROI of fewer than 100 pixels due to statistical factors, because matrices of count numbers can be under-populated (or sparse) and, thus, texture features extracted from them are unstable. In other words, small ROIs would have insufficient numbers of pixels to reliably compute the texture parameters. As different lesions may have different sizes, a minimum size that ensures the texture variation in each ROI is captured needs to be established.

Results for ROI size influence on the *variance* of the extracted feature show that the size of the ROI does affect the level of variation. The variation is inversely proportional to the size of ROI up to 900 pixels. This can be clearly seen in the case of the gradient feature. For a size of ROI larger than 900 pixels, however, the variation is relatively small and consistent for all features. As such, for any application of texture analysis technique to characterise biological tissue, a minimum ROI size has to be determined to ensure the extracted features are reliable. Due to the nature of variance, when it is used for classification, texture features extracted from a very small ROI size will affect the accuracy of the method.

The results for ROI size influence on the *mean* of the extracted feature demonstrated a clear monotonic increase in the extracted value for the RLM feature as the ROI size increases, indicating that this feature is ROI-size dependent. This finding is in agreement with the findings by Chan and McCarty (1990). In their study, they reported that two RLM parameters (grey level non-uniformity and run length non-uniformity) vary linearly with the size of the ROI. It is important to note that a standard ROI size should be used for any particular study using RLM feature to avoid misleading results due to the difference in the ROI size. This feature, however, was excluded from our image classification study.

ROI depth is another determinant of the extracted features value. According to Oosterveled *et al.* (1991), the ultrasound beam causes a depth dependence of the B-mode image texture. This can be explained by the fact that the intensity progressively decreases as the beam advances through tissues because of scattering, refraction and absorption phenomena (Lagalla and Midiri, 1998). Time gain compensation (TGC) is typically used to amplify the amplitude of echoes in order to compensate for signal attenuation on the travel path. This compensation method, however, is position-dependent, which means that different tissues in the same TGC time-range (or corresponding depth range) will be compensated by the same amount without taking into account the non-uniform beam attenuation within the body (Xiao *et al.*, 2002) which can be caused by the different intervening tissue layer (Wilhjelm *et al.*, 1998).

The attenuation on the ultrasound signal as it passes through the tissue is one of the underlying factors that affect the B-mode image texture. Morris (1988(a)) stated that correction of this variation in the raw image is not straightforward and, indeed, may not

even be possible. To overcome this problem, according to Oosterveled *et al.* (1991), some studies have confined the ROI position to a fixed depth to avoid the depth dependency. For instance, Nicholas *et al.* (1986) set the regions selected to be within 45 to 120 mm from the transducer/skin interface while Bader *et al.* (2000) reported that in the study, all tumours were located at a depth of between 2 cm and 3 cm. A fixed ROI depth approach, however, will limit the flexibility of the texture analysis technique.

Since the position of the lesion may vary, it is crucial to use texture features which are resistant to depth variation. We have evaluated the five texture features employed in this study for their robustness to ROI depth variation. The results in Table 4.3 show that GLCM and wavelet features do not exhibit a significant relationship with the ROI depth, while the other three features (RLM, gradient and AR) do exhibit a significant relationship with the ROI depth.

Based on the findings from this chapter, RLM, gradient, and AR features, which all have a significant relationship with at least one of the parameters tested (ROI size, ROI depth, and gain), were excluded for the image classification study (Chapter 7). In addition, the ROI size was kept above 900 pixels for all assessments (except in the ROI size dependency). Nevertheless, it should be noted that these parameters can still be used by making the necessary corrections to the extracted features. For this purpose, for example, the linear regression line can be used as a standard curve to correct the extracted feature. However, this approach means the extracted features are not readily useable, hence it was not considered in this study.

4.5 Summary

The experimental evaluation of five texture features presented in this chapter aims to determine the robustness of a specific texture analysis method. A tissue-equivalent phantom was used as a test object. Several key factors that may affect the performance of texture analysis were considered, namely ROI size, ROI depth, and scanner gain setting. The susceptibility of these methods to ROI size was investigated in two ways: the *variance* and the *mean* of the extracted texture features. The first was evaluated using CoV while the second was evaluated using correlation coefficient. Susceptibility to the ROI depth and the scanner gain setting were also evaluated using correlation coefficient. The ROI depth dependency was evaluated at three different scanner focuses.

Out of the five texture features studied, GLCM and wavelet were found not to be significantly affected by the factors studied and were considered for image classification (see Chapter 7). This would allow more flexibility of the texture analysis technique employed. On the other hand, RLM is affected by ROI size, ROI depth and gain; gradient is affected by ROI depth and ROI sizes of smaller than 900 pixels; and AR is affected by ROI size and ROI depth. Texture features which are affected by at least one of the factors evaluated were not used for the image classification study (Chapter 7).

Chapter 5
Repeatability and Reproducibility – Phantom Studies

The objective of this chapter is to determine the variability and reliability of image texture caused by random variation during image acquisition, which will affect the reliability of the texture analysis technique. This has been broken down into 3 parts, which are to determine:

(i) The repeatability (consistency) of the images produced due to the random variation caused by the ultrasound system.
(ii) The repeatability (consistency) of the images produced due to the random variation caused by the operator.
(iii) The reproducibility (agreement) of the images acquired by two operators under identical condition: by how much are the two likely to differ.

Section 5.1 reviews the importance of repeatability and reproducibility of the image texture and possible contributing factors. The methodology and statistical analysis are explained in Section 5.2. Section 5.3 presents the results, and these are discussed in Section 5.4. Finally, Section 5.5 summarises this chapter. The study was carried out using a tissue-equivalent phantom.

5.1 Background

The textures of the acquired images are subject to random variations. There are two factors that may contribute to these variations during image acquisition were considered in this study: the ultrasound system and the operator.

In this study, repeatability refers to the reliability of the instrument (ultrasound) and the operator in producing a consistent image (in terms of its texture) when repeated within a

short period of time and under identical condition, while reproducibility refers to the agreement between two operators performing a similar scan.

Reliability, which refers to the repeatability and reproducibility of the measurement when repeated on the same subject, determines the sensitivity and the specificity when used for classifications or predictions (Lachin, 2004). According to Li *et al.* (2004(a)) the reliability of a measurement technique may affect the accuracy of a diagnosis. In addition, according to Bailey *et al.* (2007), it is essential to show that a single operator can obtain the same results when repeated measurements are made using the same method under identical conditions. Since ultrasound is an operator-dependent modality (Collinger *et al.*, 2009), the degree of variation caused by the operator (intra-operator) also needs to be established to ensure that the application of a texture analysis technique on ultrasound images is reliable. In addition, the inter-operator reproducibility is equally important. According to Karanicolas et al. (2009), instruments that are useful in clinical or research practice will, when the object of measurement is stable, yield similar results when applied by different operators. The agreement between two measurements is particularly important when there may be underlying natural variation (Kirkwood and Sterne, 2003). In this particular case, the variation in the image acquired occurs due to the randomness in scanning carried out by the operator. Since texture analysis is able to quantify information which may not be perceived visually (Bocchi *et al.*, 1997), it is important to note that two images which look similar visually may not necessarily be texturally similar.

Shen *et al.* (1997) emphasised the importance of reliability of the measurement in quantitative methods. More specifically, for B-mode ultrasound, Chambless *et al.*

(1996) stressed on the importance of thoroughly assessing the amount of total variation and its source. Therefore, low variations in the image produced (due to the randomness in the scanner and the operator) are crucial in texture analysis to ensure that differences in the derived textural features are due to the differences in the acoustic properties of the reflecting material, not due to the deterioration of the system or process component.

5.2 Methodology and Statistical Analysis

The basis used for the experimental setup for the repeatability test was the document "Measurement Good Practice Guide No. 52" by the National Physical Laboratory (2005).

> *"The repeatability of the measurement can be quantified in either of 2 ways:*
>
> *(i) Set the equipment to perform a number of consecutive tests (e.g. 10) using identical conditions and without removing the specimen between measurements. In this case, the only variables are those relating to the performance of the measurement system and the associated statistical (random) effects in data capturing and analysis by the software. For simplicity, this will be referred to as 'instrument-only repeatability'.*
>
> *(ii) As above, but with the specimen removed completely in between measurements. In this case, effects due to variability in the test set-up are introduced. These will include resetting the device and the precision of repositioning the sample. This will be referred as the 'instrument-operator repeatability'."*

5.2.1 Repeatability of Ultrasound System

The objective of this section is to assess the reliability (consistency) of the images produced due to the random variations caused by the ultrasound system.

Ten images were acquired using a Toshiba Xario scanner with a 6.0 MHz transvaginal transducer (model PVT-661VT). An RMI phantom model 403 GS LE was used as a test object. A stand was used to hold the probe during the image acquisition. The machine setting was kept constant across image acquisition process, set to the manufacturer's pre-defined setting, *Endo-Vaginal-Gynae***.** The images were transferred to the PC in *.bmp* format for further analysis. Please refer to Figure 4.1 (Chapter 4) for an illustration of the experimental setup and the analysis process flow. A square ROI (50X50 pixels) (see Figure 4.2 in Chapter 4) was drawn in the centre of the image for features extraction using the MaZda version 4.6 texture analysis software.

A coefficient of variation (CoV) (see section 3.8.1) was calculated for each texture feature. The ultrasound repeatability COV reflects the variability of the images due to random variations in the ultrasound system. The lower the CoV, the smaller the variation between the repeats: thus, the higher the repeatability (Bailey *et al.*, 2007).

5.2.2 Repeatability of Intra-operator

The objective of this section is to assess the reliability (consistency) of the images acquired in terms of the random variations caused by the operator.

Two operators acquired a series of B-mode images (10 scans for each operator) from the test-object (RMI-403 GS LE). The transducer was removed from the test object in between scans. The images were acquired using a Toshiba Xario scanner with a 6.0 MHz transvaginal transducer (model PVT-661VT). The machine setting was kept constant across the scanning process for both operators. As in the ultrasound system repeatability method, the settings were set to the manufacturer's pre-defined setting, *Endo-Vaginal-Gynae***.** Figure 5.1 illustrates the image acquisition and the process flow of the intra-operator repeatability assessment. A square ROI (50X50 pixels) was drawn in the centre of the image for features extraction using MaZda version 4.6.

A coefficient of variation (CoV) (see section 3.8.1) was calculated for each texture feature. Intra-operator CoV reflects the variability of the images due to random variations in the scans caused by the operator.

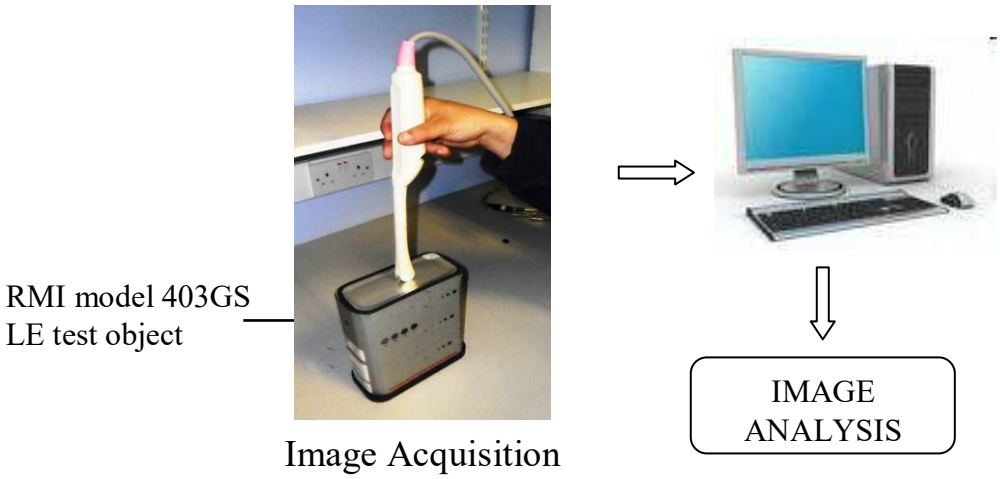


Figure 5.1: Block diagram of intra-operator repeatability assessment.

5.2.3 Reproducibility of Inter-operator

The objective of this section is to assess the degree of the difference (which reflects the agreement) of the images produced by two operators on the same object under identical circumstances.

Using the same images obtained from the intra-operator study, texture features derived from the images obtained by the two operators were evaluated visually using a Bland-Altman plot (Bland and Altman, 1986). To draw this plot, the difference between the two operators was calculated for each image pair. The differences in the results from the two operators were then plotted against the mean value. The Bland-Altman plot allows the relationship between the variance/difference and the mean of the measured value to be investigated visually (see section 3.8.5). In this way, it is possible to evaluate whether the two operators agree sufficiently. Besides using Bland-Altman Plot, CoV (see section 3.8.1) was also calculated to objectively measure the inter-operator reproducibility.

5.3 Results

5.3.1 Repeatability of Ultrasound System

Ten consecutive images were acquired from the test object. The CoVs for the five derived textural features range from 0.26% to 2.56%, with the AR feature showing the smallest variation (0.26%) and the Gradient feature showing the largest variation (2.56%). Table 5.1 summarises the mean, the standard deviation (SD), and the CoV for all the textural features evaluated in this study. The CoV is also presented visually in Figure 5.2.

Table 5.1: Ultrasound scanner repeatability.

Textural Features	Mean	SD	CoV (%)
GLCM	984	8	0.84
RLM	387	1	0.36
Gradient	253	6	2.56
AR-model	0.28925	0.00074	0.26
Wavelet	3,794	57	1.51

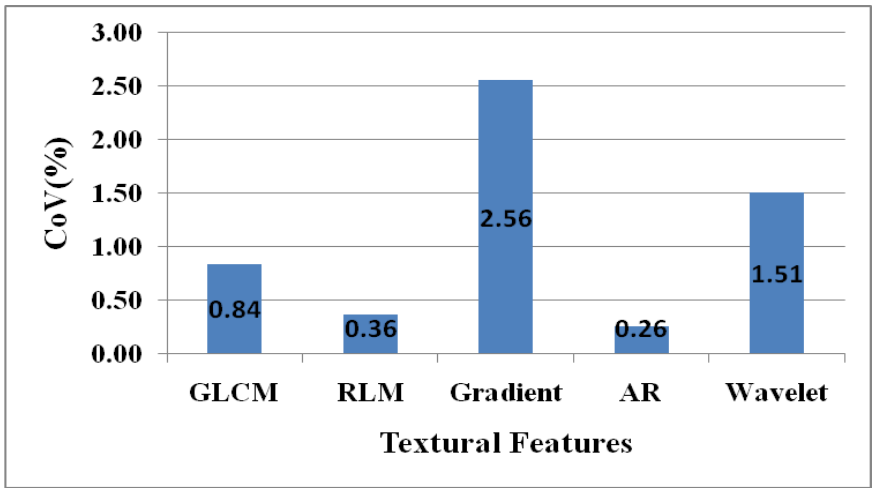


Figure 5.2: Coefficient of variation (CoV) for ultrasound system repeatability.

5.3.2 Repeatability of Intra-operator

The value of the intra-operator CoV, which demonstrates the consistency of the repeatedly acquired images, ranges from 1.04% to 5.32% for operator 1 while for operator 2, the CoV ranges from 1.21% to 5.98%; with gradient features exhibiting the largest variation for both operators. Table 5.2 summarises the CoVs for all five textural features evaluated in this study. The mean CoV is illustrated in Figure 5.3.

Table 5.2: Intra-operator repeatability.

	Features (CoV (%))				
Operator	**GLCM**	**RLM**	**Gradient**	**AR-model**	**Wavelet**
Operator1	1.80	4.44	5.32	1.04	1.67
Operator2	1.21	5.46	5.98	1.42	2.06
Mean CoV	1.51	4.95	5.65	1.23	1.87

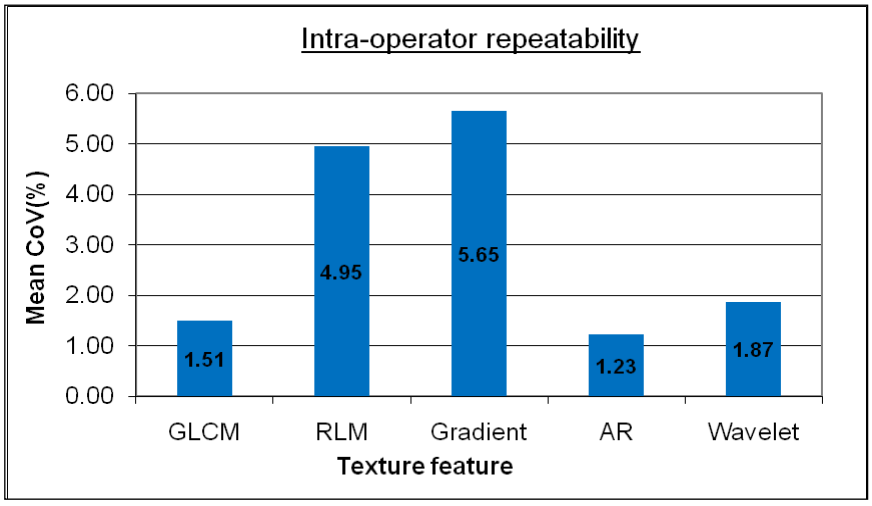


Figure 5.3: The mean CoV for intra-operator repeatability.

5.3.3 Reproducibility of Inter-operator

The inter-observer mean differences +/- SD between paired images for five textural features tested are presented in Table 5.3. Bland-Altman plots were drawn for the five features for graphical evaluation of the agreement between the textural features derived from the images obtained by the two operators. Figures 5.4(a) and 5.4(b) are examples of Bland-Altman plots showing the distribution of the GLCM feature and RLM feature respectively. All five features show similar patterns, i.e. a random variation in the difference between the two operators. The differences in the paired derived textural features between the two operators do not exhibit any systematic pattern, as the mean value increases for all textural features. In other words, the difference between the two operators does not increase (or decrease) systematically as the mean value increases.

The mean CoV for inter-operator repeatability ranges from 1.20% to 7.95%, with the GLCM feature and the RLM feature demonstrating the lowest and highest CoVs respectively. Three features (GLCM, AR and Wavelet) show very low variation, with values below 2%, while the other two features (RLM and Gradient) have values of approximately 8%. The mean CoVs for all five features are presented in Table 5.4.

Table 5.3: Inter-operator mean difference and standard deviation.

	GLCM	RLM	Gradient	AR	Wavelet
Mean Diff.	-12	32	-34	0.00107	33
Std. Dev.	18	24	27	0.00607	120

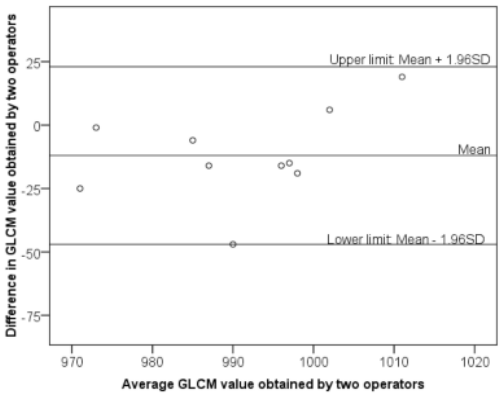


Figure 5.4(a): Example of Bland-Altman plot showing the random distribution of the difference between the two operators (GLCM feature).

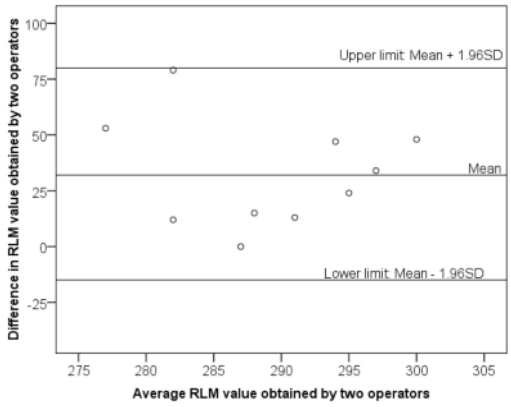


Figure 5.4(b): Example of Bland-Altman plot showing the random distribution of the difference between the two operators (RLM feature).

Table 5.4: Coefficient of variation for inter-operator reproducibility.

Mean Coefficient of Variation (CoV (%))				
GLCM	**RLM**	**Gradient**	**AR**	**Wavelet**
1.20	7.95	7.68	1.27	1.88

5.4 Discussion

Repeated measurements are likely to vary (Swinscow, 1997). The degree of the variation will determine the reliability of the measurements made. Reliability refers to the consistency or repeatability of such measurements, and is usually performed to assess the instrument's and the operator's performances (Bruton *et al.*, 2000). According to de Vet *et al.* (2006), reliability is an essential requirement for measuring outcomes in medical science, such as the assessment of radiographs.

In this chapter, a set of experiments were carried out to evaluate the reliability of B-mode images, particularly in the random variations inherited from the ultrasound system and from the operator. Since the reliability of the measurement (image acquisition) determines the sensitivity and the specificity when used for classifications or predictions, it is important to establish the degree of these variations. The scatter (variation) in the extracted texture features was evaluated as a measure to determine reliability of the image texture. According to Fry (2002), the scatter in the data is a useful measure for quantifying the effect each parameter has on the accuracy of the measurement. The variation due to image acquisition procedure can be further divided into the ultrasound system and the operator.

Generally, all the textural features evaluated demonstrated good repeatability for the ultrasound system, with all features showing less than 5% of variance from the mean. These variations, among others, may come from the electronic circuit and the element (crystal) of the transducer. The results indicate that ultrasound scanners are able to produce reliable B-mode images. A similar study by Guggenbuhl *et al.* (2008) on Computed-

Tomography (CT) scanner images reported that the COV for GLCM parameters ranged from 0.03% to 3.32%, which is consistent with our findings for ultrasound scanners.

In this chapter, we also assessed the intra- and inter-operator reliability in image acquisition. It was observed that the relative variations among those five features evaluated for ultrasound system and intra-operator exhibit a relatively consistent pattern: the AR feature has the smallest variation while the gradient feature has the largest variation. The reproducibility of both operators for all features is comparable. It was also observed for all features that their intra-operator CoVs are higher than the ultrasound system CoVs. This can be explained by the fact that the variations in the scanning process are a combination of both the ultrasound system and the operator. Nielson *et al.* (2000) carried out a repeatability assessment of the first-order descriptive statistics parameters which were extracted from B-mode image. They reported that intra-operator CoV repeatability ranges from 5% to 35%, which is higher than our findings. One of the possible reasons for this is that first-order descriptive statistics only take into account the intensity, and not the relationship between pixels, which makes this technique more susceptible to random variations.

For inter-operator reproducibility, the Bland-Altman plots clearly show that the differences between the two operators did not appear to increase as the mean value of extracted GLCM feature increases. The same pattern was observed for the four other features: RLM, AR, Gradient and Wavelet features. In other words, there is no obvious relationship noted between the differences of the two operators and their means for all five features evaluated. This indicates that increases in the derived feature values do not affect the difference between the two operators. According to Bland and Altman (2010), if the differences are proportional to the mean, a logarithmic transformation of the data

has to be performed before analysis is carried out. The CoV calculated to measure the consistency of the image produced by two different operators shows relatively small variations (good agreement).

From the results, it can be concluded that the ultrasound system is able to produce consistent images. Good agreement between two operators, due to randomness in the scanning process, is also achievable. Since the results show that scanning using a phantom is able to produce consistent images, it is worth continuing to an assessment of scanning repeatability in a clinical environment, which is relatively more complex. This is addressed in the next chapter (Chapter 6).

5.5 Summary

The consistency of the image texture is an important factor in texture analysis. The repeatability and reproducibility issues have therefore been assessed to determine the consistency in the image texture, which could be contributed to by random variations inherent in the ultrasound scanner and the scanning process. A tissue-equivalent phantom was used for this purpose. This chapter also evaluated the agreement in the images produced by two different operators. The results for the ultrasound system repeatability show low variation in the images produced under identical conditions. The intra-operator repeatability also demonstrated relatively low variations and good agreement was found in the images obtained by two different operators for all textural features.

Based on the findings from this chapter, it can be concluded that ultrasound systems are able to produce consistent images. This chapter has also established that intra-operator repeatability and good agreement between two operators are achievable. Therefore, it is worth continuing to a similar assessment in a clinical setting, which involves more complex scanning processes (Chapter 6).

Chapter 6
Reliability of ROI Segmentation and Clinical Image Acquisition

The objective of this chapter is to assess the factors related to clinical images that might potentially affect the reliability of the extracted texture features, namely ROI segmentation, clinical image acquisition, and 'calliper line'. Section 6.1 reviews the influence of those three factors on the extracted features. The methodology and statistical analysis are explained in Section 6.2. Section 6.3 presents the results, and these are discussed in Section 6.4. Finally, Section 6.5 summarises this chapter.

6.1 Background

According to Menke *et al.* (2009), the reliability of any diagnostic test is important for diagnostic precision. In particular, the reproducibility of B-mode images is important if the images are to be used for texture analysis. In order to assess the reliability of clinical images, we first need to assess the reliability of the ROI segmentation, i.e. the process of delineating the boundaries of the structure of interest.

Segmentation is an important step in ultrasound B-mode image analysis (Xiao *et al.*, 2002). To ensure the reliability of the extracted features, it is necessary to accurately identify and segment the area of interest, and the segmentation procedure should be repeatable. Image noise and distortions that occur in ultrasound images, however, pose a challenge in the ROI segmentation of ultrasound images (Booth *et al.*, 2006), potentially resulting in missing boundaries (Noble & Boukerroui, 2006). For instance, Coleman *et al.* (2005) reported that the delineation of the boundaries of the region of interest is more difficult in vivo images than in vitro images.

As discussed in Chapter 5, this variability will influence the accuracy of the classification process. Although the repeatability of B-mode images was established in Chapter 5, it is important to note that the process of obtaining a B-mode image from a phantom is relatively easier than from an actual scanning process in a clinical environment. Therefore, the reliability (repeatability) of B-mode image acquisition in a clinical setting also needs to be established.

As mentioned earlier, to establish the reliability of clinical images, the reliability of the ROI segmentation must first be assessed. The purpose of image segmentation is to divide an image into regions which are meaningful for a particular task (Dougherty, 2009). According to Nailon (2010), a significant inter- and intra-clinician variability has been reported in the segmenting of tumours of the lung, prostate, brain and oesophagus, and this variability has been shown to be correlated with the digital imaging modality used. Ultrasound images are arguably the most difficult among various imaging modalities to perform segmentation upon (Noble, 2010).

There are many studies that have applied automatic segmentation for quantitative analysis (for example, Lugt *et al.*, 1998; Boukerroui *et al.*, 1998; Rusko *et al.*, 2009; Kakar and Olsen, 2009). However, it is important to note that ultrasound image segmentation is strongly influenced by the quality of the image (Noble & Boukerroui, 2006). According to Rocha *et al.* (2011), ultrasound images are noisy and large parts of the edges may be missing, producing gaps in organ boundaries which pose a great challenge to automatic segmentation. Similarly, Liu *et al.* (2010) stated that segmentation of breast ultrasound image is very difficult due to blurry boundaries.

Two steps can be applied to improve the ROI segmentation reliability: (i) by introducing image enhancement, and (ii) by applying a semi-automatic segmentation technique. Image enhancement is the processing of images to improve their appearance to human vision (Dougherty, 2009; Cheng *et al.,* 2010), in terms of better contrast and visibility of features of interest (Dougherty, 2009), while semi-automatic segmentation, which involves manual segmentation using dedicated tools such as *magnetic lasso tool*, will aid in the outlining step. The use of *magnetic lasso tool* for the segmentation of medical images has been proposed by Park *et al.* (2005).

Another issue that needs to be considered, particularly with clinical B-mode images, is the 'calliper line'. The images used for classification in this study were obtained from an existing database. Since the images were not originally acquired for texture analysis purposes, there is a 'calliper line' in the image, which is a line that is used to measure the size of the lesion under 'conventional' methods of interpreting B-mode images. Using a texture feature that is not affected by the 'calliper line' means the method is more flexible to be applied to B-mode images.

6.2 Methodology and Statistical Analysis

To assess ROI segmentation and clinical image reliability, Adobe Photoshop 6.0 (Adobe Photoshop 6.0, Adobe System Incorporated) and MaZda 4.6 (Institute of Electronics, Technical University of Lodz, Poland) were used. For both studies, the process involved can be divided into two main steps: (i) ROI segmentation using Adobe

Photoshop 6.0, and (ii) features extraction using MaZda 4.6. These steps are illustrated in Figure 6.1.

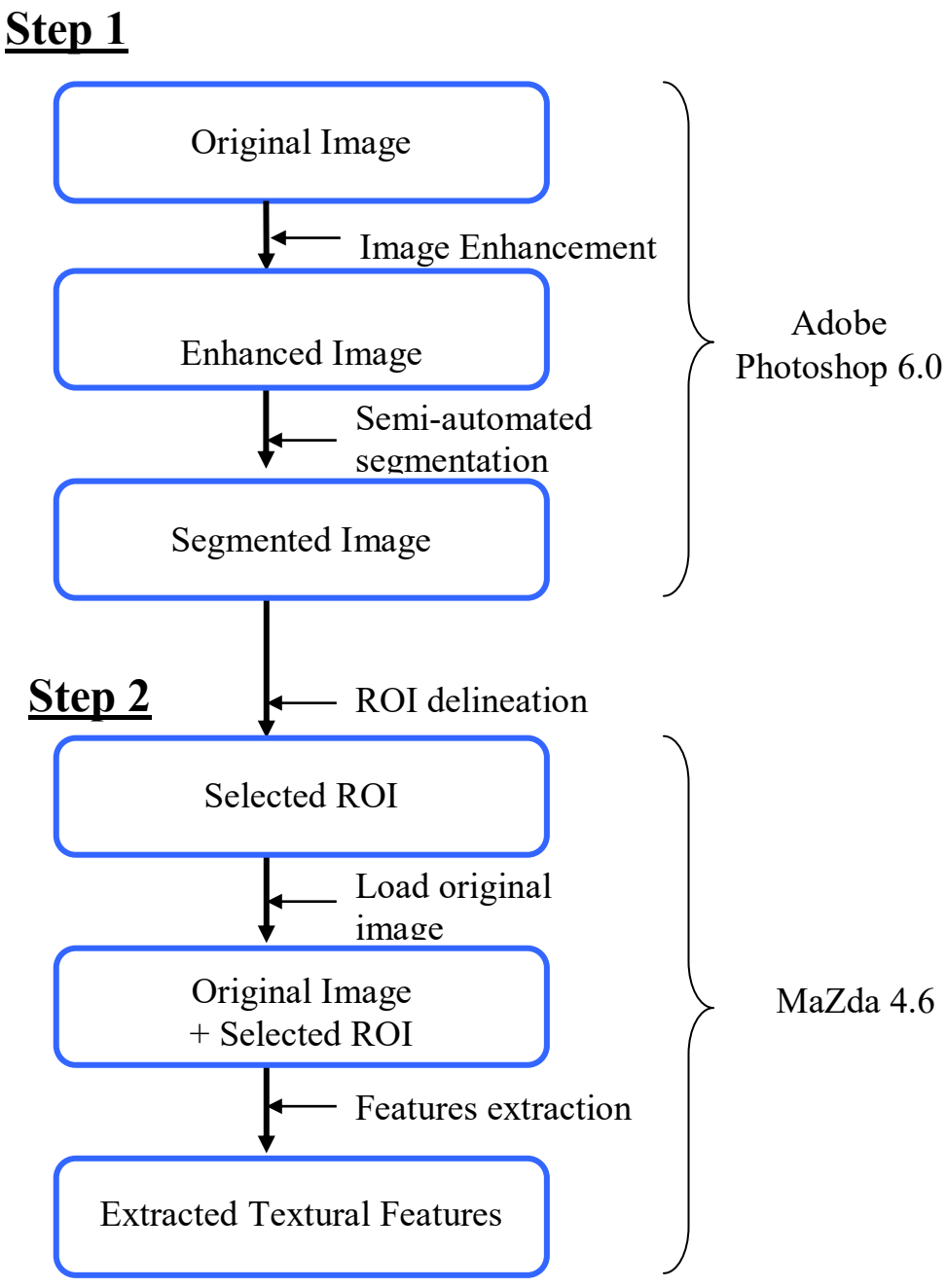


Figure 6.1: Flow-chart of ROI segmentation and clinical image repeatability assessment.

6.2.1 Repeatability of ROI Segmentation

The objective of this section is to establish the repeatability of the ROI segmentation procedure.

In order to assess the repeatability of the ROI segmentation procedure, ten historically representative images were used. The images were transferred into personal computer (PC) for analysis, i.e. the analysis was performed off-line. The original images were first pre-processed prior to image segmentation to enhance the image. According to Drukker *et al.* (2003), image enhancement is an important step before segmentation is carried out as medical ultrasound images include such considerable noise that segmentation becomes difficult. Image enhancement helps in segmentation by suppressing image noise, and enhances the contrast between the lesion area and the tissue background (Shi *et al.*, 2010).

The lesion contour was first outlined by an ultrasonologist for ROI segmentation. The ROI was chosen to be as large as possible in order to cover the whole lesion (see Figure 6.2(c)). Using Adobe Photoshop, contrast enhancement was performed by applying a *gradient-filter* to the original image (Figure 6.2(a)). The enhanced image clearly shows an image with better contrast and sharper edges (Figure 6.2(b)). Based on the contour drawn by the ultrasonologist, the area of interest was manually segmented using the *magnetic lasso tool* (Figure 6.2(c)) which yields a temporary segmented image (Figure 6.2(d)). The use of the *magnetic lasso tool* has been suggested by Park *et al.* (2005) for the segmentation of the anatomical structures in medical images on a personal computer. Based on the temporary segmented image, MaZda version 4.6 was the used to outline the ROI (Figure 6.2(e)) which results in the selected ROI (Figure 6.2(f)). Finally, the

original image was loaded (Figure 6.2(g)) followed by the features extraction. Each image was delineated 10 times. The delineation was performed for each image once before being repeated for the subsequent delineation. Figure 6.2 illustrates the clinical image segmentation assessment.

To assess the level of ROI segmentation repeatability, i.e. the variation in the values of the extracted features due to the ROI segmentation procedure, CoV (see Section 3.8.1) was calculated for each image set. The CoVs were then averaged for the ten image sets to reflect the overall repeatability. This CoV reflects the repeatability of the ROI segmentation procedure.

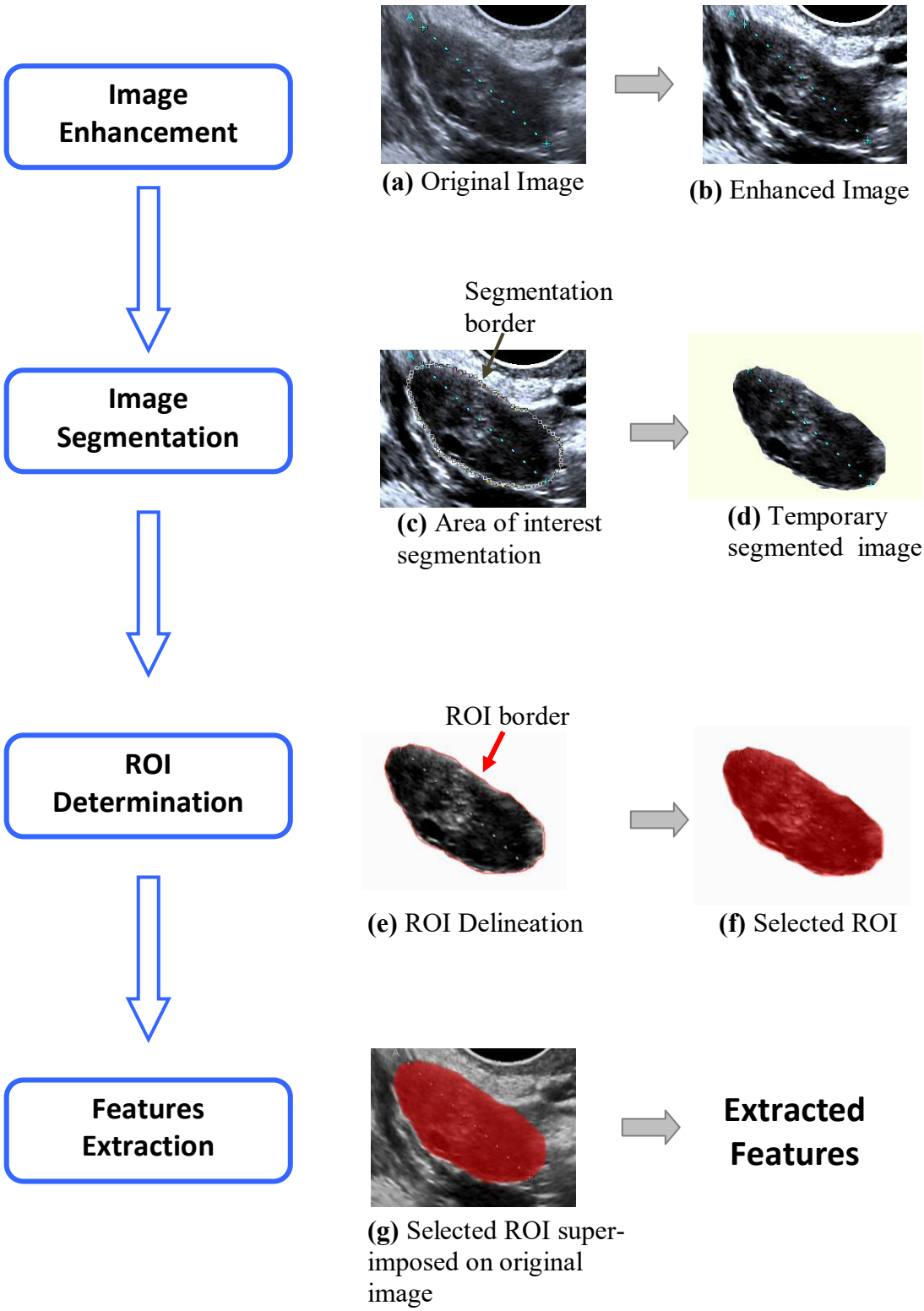
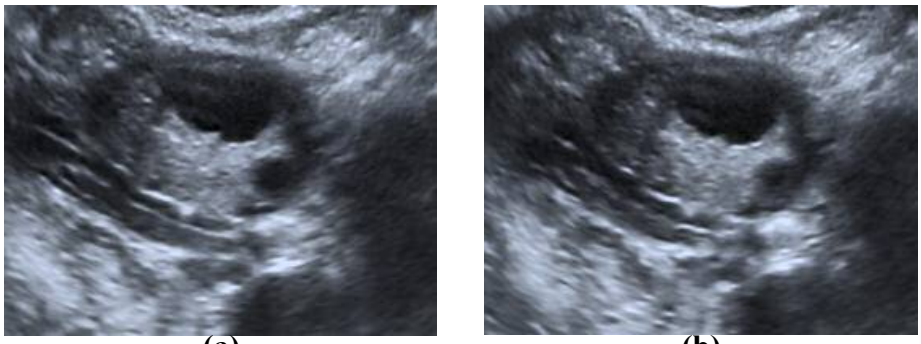


Figure 6.2: Clinical image segmentation assessment diagram.

6.2.2 Repeatability of Clinical Image Acquisition

The objective of this section is to assess the reliability (consistency) of the clinical images acquired (in terms of texture) due to the random variations caused by the operator during image acquisition.

In order to assess the repeatability of the B-mode image acquisition in clinical practice, particularly for ovarian images, five sets of images were used in this study. Each image set consisted of ten images acquired by a trained operator. The features extraction procedures for clinical image repeatability assessment are similar to the procedures for ROI delineations repeatability assessment (see Section 6.2.1) except that each image was only delineated once. Figures 6.3(a) and 6.3(b) are examples of two clinical images belonging to the same image set (group).



(a) **(b)**

Figure 6.3: Example of two images acquired from the same subject.

To assess the level of the clinical image acquisition repeatability, i.e. the variation in the clinical image texture due to the scanning process, CoV (see Section 3.8.1) was calculated for each set of images. The CoVs were averaged from all five sets of images to reflect the overall repeatability. This mean COV reflects the reliability of the image acquisition procedure in a clinical environment.

6.2.3 Influence of Calliper Line

The objective of this section is to determine the effect of a calliper line (which is normally used in clinical images) on the extracted texture features.

In order to assess the influence of calliper line, ten images were used. The original image (Figure 6.4(a)) was first opened using Adobe Photoshop. This image was then saved to a new file without imposing any changes (Figure 6.4(b)). The simulated calliper line was then drawn in the middle of the image and saved to another new file (Figure 6.4(c)). To ensure that the colour of the simulated calliper line was identical to the actual calliper line, a feature in Adobe Photoshop called 'colour picker' was used. The same procedures were carried out on all the images. Finally, texture features were extracted from these two new sets of images.

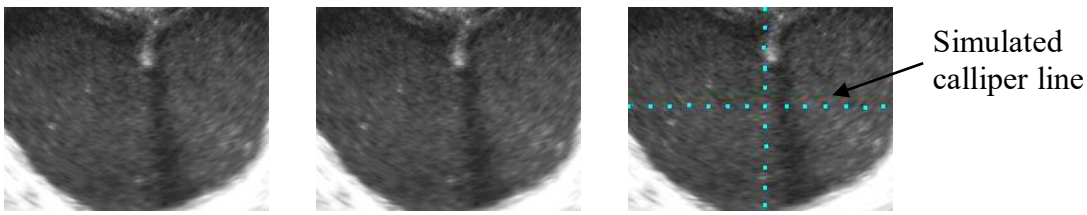


Figure 6.4: Illustration of calliper line assessment: (a) Original image; (b) Image saved using Adobe Photoshop (group1); and (c) Image with simulated 'calliper line' (group2).

A Q-Q plot was used to determine the normality of the data. For a normal distribution, the plot should show a linear relationship. Figure 6.5 shows an example of the normally distributed data (RLM feature of Group 1) while Figure 6.6 shows an example of the non-normally distributed data (GLCM feature of Group 1). The two-sample t-test (for the normally distributed data) and the Wilcoxon-Mann-Whitney- test (for the non-normally distributed data) were used to determine the difference in the texture features values extracted from the two groups, i.e. the images with and without simulated calliper lines. The statistical analysis was performed using SPSS version 16.0 for Windows (SPSS Inc., Chicago, Illinois, USA).

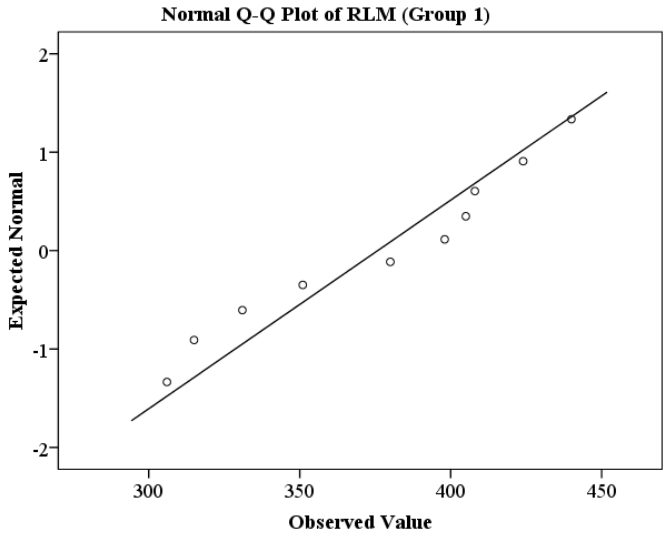


Figure 6.5 – Example of Q-Q plot for normally distributed data (RLM feature of Group1).

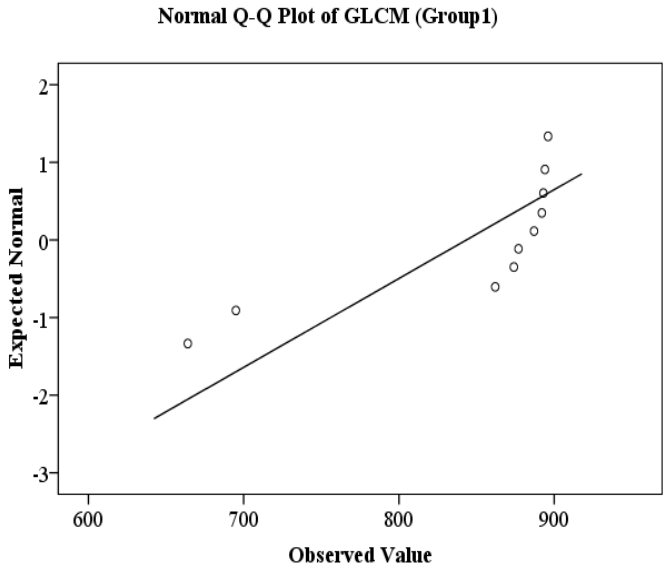


Figure 6.6 - Example of Q-Q plot for non-normally distributed data (GLCM feature of Group1).

6.3 Results

6.3.1 ROI Segmentation Repeatability

For ROI segmentation repeatability, the CoV range for GLCM was 0.40% to 1.95%; for RLM it was 0.60% to 4.62%; for gradient, 0.39% to 2.50%; for AR, 0.02% to 0.44%; and for wavelet, 0.02% to 1.60%. The mean CoV for all features evaluated ranged from 0.17% to 1.90% with AR exhibiting the smallest variation (0.17%) and RLM exhibiting the highest variation (1.90%). The CoV for all image sets and their means and are shown in Table 6.1. The mean CoV for all five features are graphically shown in Figure 6.7.

Table 6.1: Coefficient of variation (CoV) for ROI segmentation repeatability.

	Coefficient of Variation (%)				
Image	**GLCM**	**RLM**	**Gradient**	**AR**	**Wavelet**
1	1.93	1.16	2.50	0.44	1.25
2	1.95	0.60	2.29	0.31	1.14
3	0.76	4.62	2.26	0.16	1.08
4	0.95	2.87	1.94	0.30	1.23
5	0.47	1.56	1.39	0.18	0.55
6	0.65	1.16	1.12	0.12	0.75
7	0.73	2.10	0.89	0.06	1.42
8	0.50	2.20	0.77	0.08	1.60
9	0.66	1.43	0.60	0.04	0.34
10	0.40	1.27	0.39	0.02	0.02
Mean CoV (%)	0.90	1.90	1.42	0.17	0.98

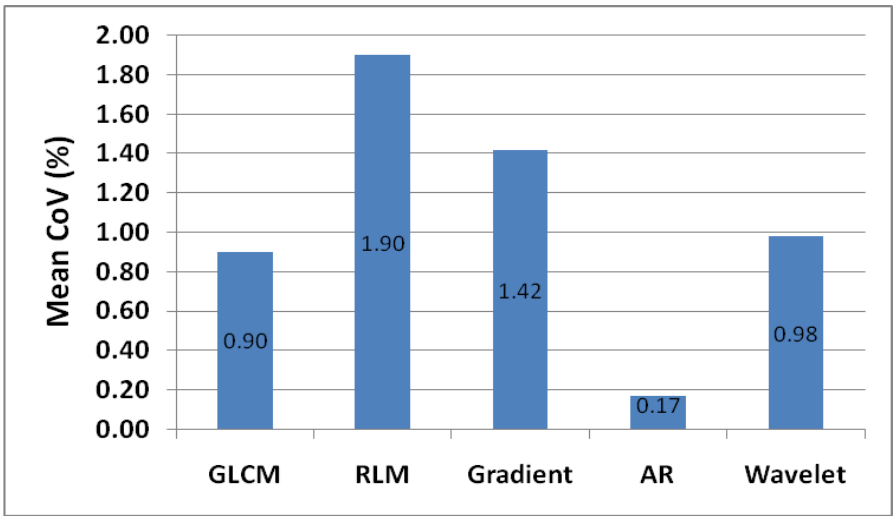


Figure 6.7: Graph showing the mean CoV for the ROI segmentation repeatability.

6.3.2 Clinical Image Repeatability

For clinical image repeatability, the CoV values which reflect the variability of the extracted features for GLCM ranged from 0.69% to 3.35%; for RLM, 1.33% to 5.24%; for gradient, 4.61% to 11.83%; for AR, 0.27% to 0.64%; and for wavelet, 0.84% to 1.99%. The mean CoV for all five features evaluated ranged from 0.50% to 7.77%, with AR having the smallest variations and gradient having the largest variations. The CoV for all images and their means and are shown in Table 6.2. Figure 6.8 shows the mean CoV for all five features included in this study.

Table 6.2: Coefficient of variation (CoV) for clinical image repeatability.

	Coefficient of Variations (%)				
Image	**GLCM**	**RLM**	**Gradient**	**AR**	**Wavelet**
Set 1	1.52	1.33	5.49	0.64	1.99
Set 2	0.69	3.89	11.83	0.27	1.28
Set 3	3.35	3.37	4.61	0.42	0.84
Set 4	1.97	3.78	6.14	0.60	1.49
Set 5	3.20	5.24	10.80	0.56	1.01
Mean CoV (%)	2.15	3.52	7.77	0.50	1.32

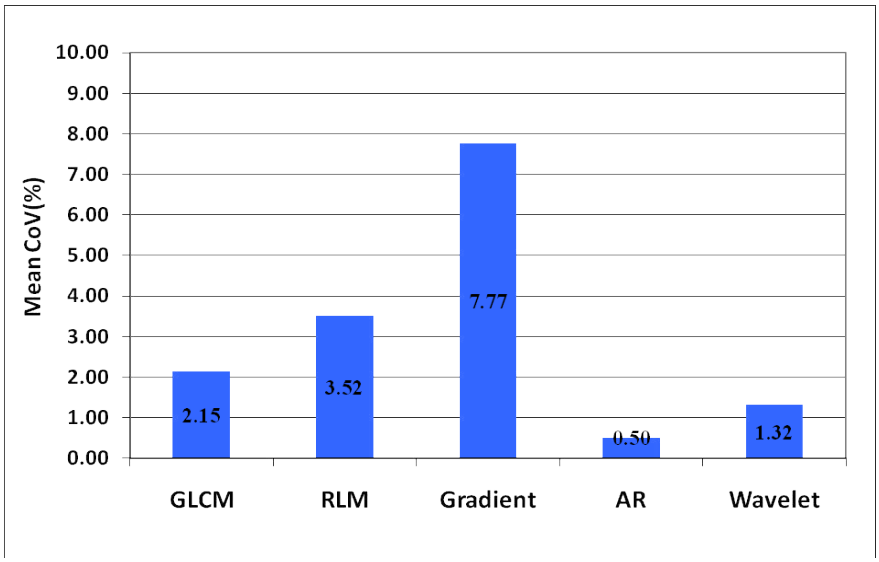


Figure 6.8: Graph showing the mean CoV for the clinical image repeatability.

6.3.3 Influence of Calliper Line

The results show that three features do not have a significant difference between images with and without calliper line. Those features are GLCM (p=0.29), RLM (p=0.71), and wavelet (p=0.43). The other two features, (gradient and AR), however, show a significant difference with p-values of less than 0.01. The mean, median, standard error of the mean (s.e.), and p-value for all features are summarised in Table 6.3.

Table 6.3: Summary results of the calliper line assessment.

Texture Features	Image	Mean	Median	S.E.	p-value
GLCM	Group1	843	882	28	0.29
	Group2	858	868	10	
RLM	Group1	376	389	15	0.71
	Group2	384	398	16	
Gradient	Group1	151	134	16	<0.01
	Group2	60	62	5	
AR	Group1	0.24756	0.24285	0.00484	<0.01
	Group2	0.22034	0.22001	0.00083	
Wavelet	Group1	4406	4,484	66	0.43
	Group2	4478	4,529	50	

Group1=Images without calliper line; Group2=Images with calliper line.

6.4 Discussion

Any texture analysis requires segmentation of the area of interest. The purpose of segmentation is to divide an image into regions that are meaningful for a particular task (Dougherty, 2009). It should be noted that the performance of ultrasound tissue characterisation is related to the ultrasound image segmentation problem (Noble, 2010). According to Coleman (2007), to determine whether a system would be useful, the segmentation of the tissue of interest from the surrounding tissue has to be demonstrated. Nonetheless, medical image segmentation remains a common problem for all medical imaging applications (Olabarriaga and Smeulders, 2001; Zhu *et al.*, 2006; Jiang *et al.*, 2010). Ultrasound image segmentation, for example, is strongly influenced by the quality of the image (Noble and Boukerroui, 2006). There are several factors that can degrade image quality, for instance, non-uniform beam attenuation in the body can cause variations in the intensities in-homogeneities of B-mode images, which in turn may impair the segmentation process (Wilhjelm *et al.*, 1998; Xiao *et al.*, 2002). Therefore, to assess clinical image reliability, the ROI segmentation repeatability must first be established.

Munzenmayer *et al.* (2009) has applied image processing technique to remove the artifacts from the images before texture features are extracted. In this study, an image processing technique (gradient-filter) was first applied to the original images to enhance the contrast and to sharpen the edges of the lesion before the segmentation process was carried out manually using the *magnetic lasso tool*. The underlying principle of the enhancement is to enlarge the intensity difference between objects and surroundings (Shi *et al.*, 2010). Although there are automatic segmentation techniques available, as discussed in Section 6.1, these techniques are not recommended for application on

ultrasound images. Speckle, echo shadows, attenuation, and low contrast are among the factors that affect the B-mode image quality (Rocha *et al.*, 2011). In fact, Park *et al.* (2005) stated that it is not possible to automatically perform segmentation because most anatomical structures in medical images can not be identified by a computer; which can only be identified by medical experts. For instance, in a study on thyroid glands by Catherine *et al.* (2006), they used a rectangular ROI enclosing the thyroid because automatic segmentation of the whole gland was too complex. Although a rectangular ROI is easy to draw, it also has the disadvantage that it may not cover the whole lesion.

The results of this study demonstrate that the use of the *magnetic lasso tool* for semi-automatic segmentation was able to produce good segmentation repeatability. According to Hwang *et al.* (2005), using Adobe Photoshop means any anatomical structures, including pathological lesions in various types of medical images, can be segmented without difficulty. Park *et al.* (2005) predicted that the use of the *magnetic lasso tool* will be widely used for the segmentation of anatomical structures in various medical images. Besides improvement in the repeatability, the use of semi-automatic segmentation tools will also reduce the time required to outline the area of interest.

For clinical image acquisition assessment, it was observed for all features that there are variations in CoV values between different subjects. For instance, the CoV for GLCM features ranges from 0.69% to 3.02%. These variations are due to the different types of lesion. In other words, the 'complexity' of the lesion will determine the level of repeatability in each subject. The more complex a structure is, the higher the variation will be. It was also observed that the results for clinical image repeatability exhibit a similar pattern compared to the results for the phantom study (see Section 5.3.2): the

AR feature has the smallest variations while the gradient feature has the highest variations. The level of variation for all features is comparable between phantom and clinical studies. Although we previously mentioned that the scanning process in a clinical environment is complex, which means higher variations are expected in clinical image acquisition, it should be noted that the images for the clinical image repeatability assessment were acquired by an experienced operator. In fact, the results of this study prove that low variation in clinical image texture is achievable.

This chapter also assessed the influence of the calliper line which may be found on clinical images. The results show that there is no significant difference between the two groups for three of the features (GLCM, RLM, and wavelet) while the other two features (gradient and AR) were found to be affected by the calliper line.

Although all features have good ROI segmentation and clinical image acquisition repeatability, the gradient and AR features are affected by the calliper line. Therefore, the gradient and AR features were excluded from the image classification study (Chapter 7).

6.5 Summary

The consistency of image texture in clinical images is an important factor that determines the accuracy of texture analysis technique in medical image classification. In addition, the reliability of ROI segmentation is equally important because it has a direct effect on the extracted features. Therefore, the repeatability of B-mode images and the ROI segmentation of ovarian tissue have been evaluated in this chapter. Since the ROI segmentation will determine the reliability of the extracted features, an assessment has to be carried out before analysis of clinical image repeatability can be performed. In this study, the ROI was semi-automatically segmented using the *magnetic lasso tool* in Adobe Photoshop. The results proved that the *magnetic lasso tool* is a useful aid in ROI segmentation and good repeatability was achieved for all five features evaluated. Similarly, the B-mode clinical image repeatability also found to be satisfactory.

This chapter also evaluated the influence of the calliper line. The results show that the extracted texture features from images with and without calliper line are significantly different for gradient and AR features. Therefore, these features were not used for the image classification study (Chapter 7).

Chapter 7
Image Characterisation and Classification

The objective of this chapter is to investigate the ability of GLCM and Wavelet features to characterise and discriminate three types of ovarian tissue: normal tissue, cysts, and malignant tissue. Section 7.1 briefly describes the quantitative image analysis and the ROI delineation approaches employed in this chapter. The methodology and statistical analysis are explained in section 7.2. Section 7.3 presents the results, which are discussed in section 7.4. Finally, section 7.5 summarises this chapter.

7.1 Background

Ovarian masses are mainly discovered or first explored with transvaginal ultrasound (Lucidarme *et al.*, 2010). The goals of the ultrasound examination, amongst others, are the early detection of ovarian carcinoma and the differentiation of ovarian masses with regards to their clinical significance (Merz, 2007). According to Zimmer *et al.* (2003), a prompt diagnosis of ovarian malignancies is important due to the high mortality rate and the difficulty to detect a tumour in its early stages.

In spite of improvements to the ultrasound equipment, such as improvements in resolution, the interpretation of ultrasound images is still subjective and dependent on the operator experience (Shung, 2006; Huang *et al.*, 2008; Rocha *et al.*, 2011). As a result, a wide range of B-mode ultrasound performance in ovarian cancer diagnosis has been reported. For instance, Kinkel *et al.* (2000) reported the accuracy of B-mode ultrasound ranging from 65%-94%.

In order to eliminate the operator dependency and improve the diagnostic accuracy, computer-aided diagnosis (CAD) has been extensively explored. Chan *et al.* (1995) raised an important point regarding CAD, i.e. the use of a computer in the radiological image interpretation:

> *'Although a computer program may never be able to achieve the level of knowledge and cognitive capability of a radiologist, a trained computer program can perform certain tasks reproducibly and consistently without the inter-observer and intra-observer variations that are commonly observed among human observers. The ability of a CAD scheme can therefore be complementary to that of a radiologist'.*

A B-mode image diagnosis is based on a visual interpretation of the morphological features of the lesions. According to Stoitsis *et al.* (2006(b)), interpretation of medical images is often limited due to the non-systematic search patterns of human. Moreover, the human eye has limitations and is not able to distinguish between more than 15-20 shades of grey levels in an ultrasound image (Lagalla & Midiri, 1998). In addition, the prevalence of malignancy may differ between women with symptomatic and asymptomatic masses, which may in turn affect the positive and negative predictive value of a test, and, potentially, sensitivity and specificity as well (AHRQ, 2006).

On the other hand, texture analysis is more sensitive than visual analysis as it is based on the smallest element of the image (Herlidou *et al.*, 1999). Consequently, it is able to provide information that cannot be easily detected by the human eye (Herlidou *et al.*, 1999). In the medical domain, texture analysis has been explored to improve the performance of radiological tests (Van Ginneken *et al.*, 2002; Li *et al.*, 2004(b); Gruszauskas *et al.*, 2008). As discussed in chapter 2, this technique has been applied to characterise various types of tissues and diseases and promising results were reported.

The ultimate goal of texture analysis is to capture distinctive texture features that will maximise the discrimination of the analysed images. In order to achieve this, two main objectives were set in this chapter. The first objective is to characterise the distribution of texture features values for each tissue type, while the second objective is to determine their discriminatory power/ability.

In order to improve the discriminatory capabilities of the texture features, several factors that might potentially affect the reliability of the extracted values have to be considered. This is to ensure that the variations in the captured extracted features are due to the difference in the tissue structure while the variations due to other factors are kept to a minimum. The influence of these factors has been discussed in detailed and has been investigated in chapter 4, chapter 5, and chapter 6 accordingly.

Out of the five texture features included in this study (GLCM, RLM, gradient, AR, and wavelet), two features (GLCM and wavelet) were found to be robust to the variations of the ROI size, ROI depth, and gain setting. These two features also demonstrate good instrument and operator repeatability/reproducibility. Additionally, they are not influenced by the calliper line. Therefore, these two features were considered for the characterisation of the ovarian B-mode images in this study. The other three texture features (RLM, gradient, and AR), which were found to be affected by at least one of the factors evaluated, have been excluded.

In this chapter, we investigate the ability of these two features to characterise and discriminate three types of ovarian tissue and lesions; namely, normal tissue, cysts, and

malignant tissue. Examples of B-mode images of the normal tissue, cysts, and malignant tissue are shown in Figures 7.1, 7.2, and 7.3 respectively.

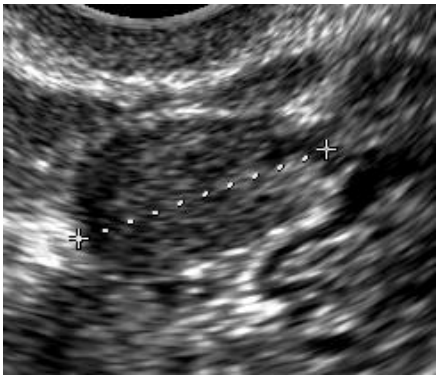


Figure 7.1: Example of normal ovary.

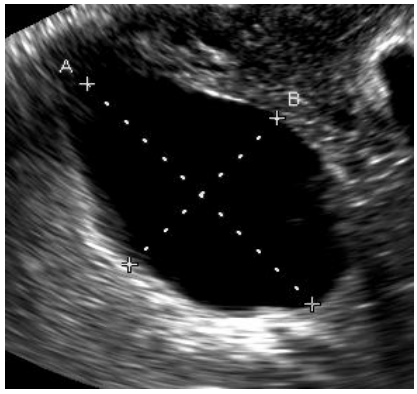


Figure 7.2: Example of cysts.

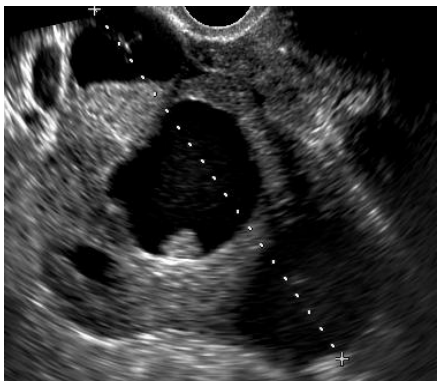


Figure 7.3: Example of malignant tissue.

Two approaches of ROI delineation were employed. For the first approach, the whole area of the tissue of interest was included. For the second approach, the anechoic area within the normal and malignant tissues was excluded.

Results from texture analysis technique can be combined with a decision-making algorithm to form a computer-aided diagnosis (CAD) system (Tourassi, 1999; Doi, 2007; Giger *et al.*, 2008) to improve the overall performance of radiological diagnosis. According to Tourasi (1999), study findings showed that CAD can enhance the diagnostic performance of a radiologist when used as a second opinion.

7.2 Methodology and Statistical Analysis

Thirty historical images consisting of normal tissue (10 images), cysts (9 images), and malignant tissue (11 images) were used in this study. The images were transferred to a personal computer (PC) for analysis, i.e. the analysis was performed off-line.

Two texture features (GLCM and wavelet) were extracted from each image. These features have been proved to be robust and reproducible, and were deemed to be suitable for B-mode quantitative image analysis. The procedures for the features extraction employed in this chapter were similar to the procedures used for the ROI segmentation and clinical image repeatability (see section 6.2.1 and section 6.2.2 for the details). In brief, Adobe Photoshop 6.0 (Adobe System Incorporated) was used to enhance the image and aid the segmentation procedure, while Mazda 4.6 (Institute of Electronics, Technical University of Lodz, Poland) was used for features extraction. The statistical analysis (Q-Q plot, t-test, Wilcoxon-Mann-Whitney (WMW) test, and ROC curve analysis) was performed using SPSS 16.0 for Windows (SPSS Inc, Chicago, Illinois, USA).

(i) Comparing Group Means

The extracted features were categorised into three groups according to the type of lesion: normal tissue, cysts, and malignant tissue. A Q-Q plot was used to determine the normality of the data. For normally distributed data, the plot should show a linear relationship. Figure 7.4 shows an example of a normally distributed data (GLCM feature of normal tissue - Approach 1) while Figure 7.5 shows an example of a non-normally distributed data (wavelet feature of malignant tissue - Approach 1).

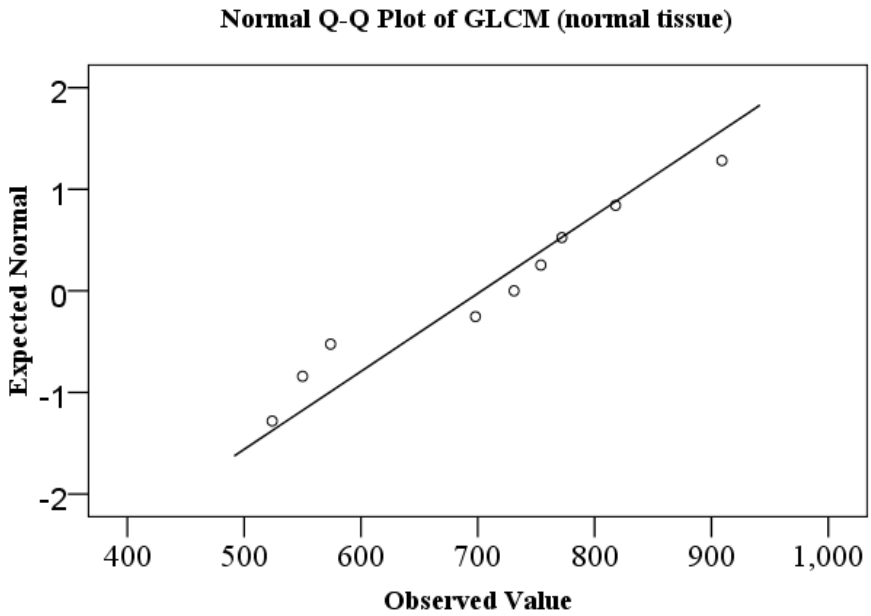


Figure 7.4: Example of Q-Q plot for a normally distributed data (GLCM feature of normal tissue - Approach 1).

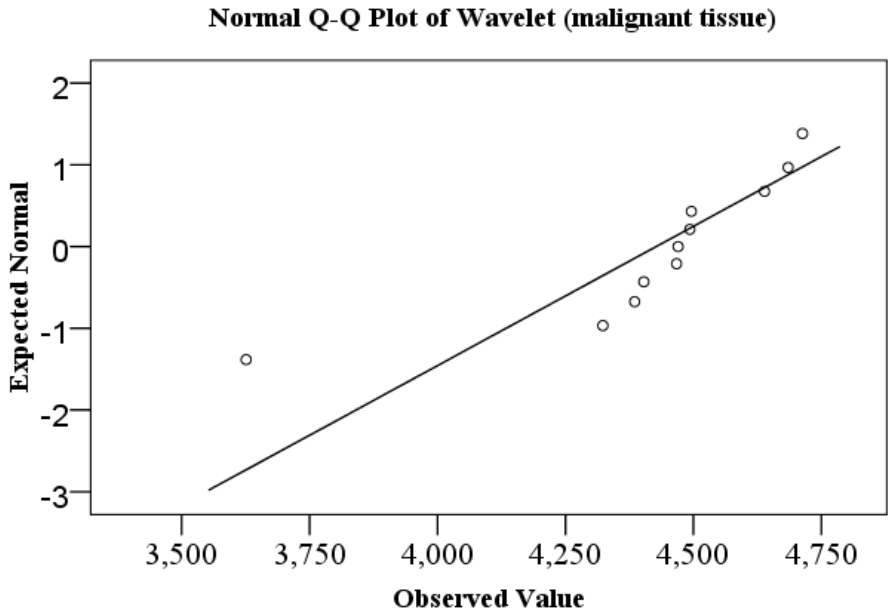


Figure 7.5: Example of Q-Q plot for a non-normally distributed data (wavelet feature of malignant tissue -Approach 1).

The two-sample t-test (for the normally distributed data) and the WMW test (for the non-normally distributed data) were used (see section 3.8.2) to determine the difference in the texture features values extracted from the three groups, i.e. to see whether or not there is a significant difference between them. The values were compared in pairs: normal tissue and cyst; normal tissue and malignant tissue; and cyst and malignant tissue. All different group pairs and ROI delineation approaches were evaluated in this manner.

(ii) Performance Assessment

A receiver operating characteristic (ROC) curve analysis (see section 3.8.4) was carried out to evaluate the ability of the texture features to discriminate between the three groups. This analysis was carried out for the group pairs that have a significant difference (p-value less than 0.05). The ROC curve analysis provides the selection of the cut-off value and accounts for the inter-dependence between sensitivity and specificity. The higher the ROC curve area value, the better the discriminatory ability. Visual inspection of the ROC curve allows us to draw an arbitrary cut-off value, which is a trade-off between sensitivity and specificity.

7.2.1 Approach 1 - Anechoic Area Included

The contour of the tissue of interest was first outlined by an ultrasonologist. Based on the contour drawn by the ultrasonologist, the ROI for features extraction was then manually segmented. The ROI was chosen to be as large as possible in order to cover the whole area of the tissue of interest.

For this approach, the anechoic area (see Figure 7.6(a)) that occurs within the ROI was also included (see Figure 7.6(b)). The texture features were then extracted using Mazda version 4.6 (Institute of Electronics, Technical University of Lodz, Poland).

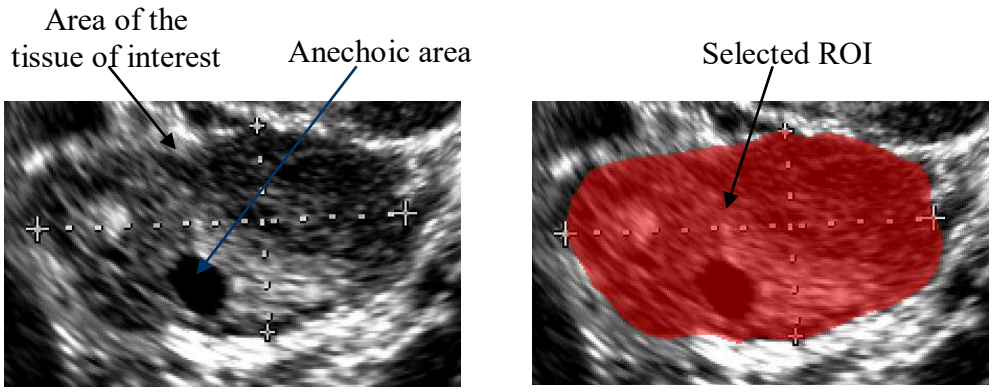


Figure 7.6: (a) Example of original image; (b) Example of selected ROI which comprises of whole area of interest.

7.2.2 Approach 2 - Anechoic Area Excluded

The procedures of outlining the ROI were similar to the procedures employed in section 7.2.1. The contour of the tissue of interest was first outlined by an ultrasonologist. Based on the contour drawn by the ultrasonologist, the ROI for features extraction was then manually segmented. The ROI was chosen to be as large as possible in order to cover the whole area of the tissue of interest

However, for this approach, the anechoic area within the normal and malignant tissues was excluded. Figure 7.7(a) and 7.7(b) illustrate the ROI determination, i.e. the texture features were only extracted from the solid area of the tissue of interest (see Figure 7.7(b)).

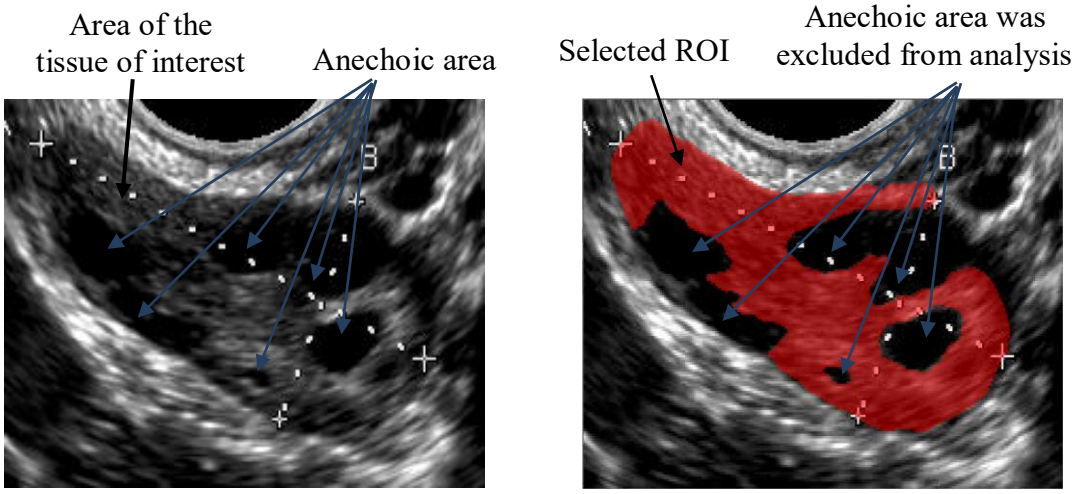


Figure 7.7: (a) Example of original image; (b) Example of selected ROI which excludes the black area occurring within the area of interest.

7.3 Results

7.3.1 ROI Delineation - Approach 1 (Anechoic Area Included)

For each group (normal tissue, cysts, and malignant tissue), the mean, median, standard deviation, and standard error of the mean (s.e.) of the extracted texture features were calculated and are summarised in Table 7.1. The mean values for the GLCM and wavelet features are visually presented in Figure 7.8 and Figure 7.9, respectively.

Table 7.1: Summary of the mean, median, standard deviation, and std. error of mean of the extracted features (ROI delineation - approach 1).

Texture Feature	Tissue Type	Mean	Median	SD	S.E.
GLCM	Normal	703	731	130	43
	Cysts	435	421	163	51
	Malignant	782	782	64	19
Wavelet	Normal	4,481	4,500	191	63
	Cysts	4,705	4,622	316	100
	Malignant	4,427	4,470	293	88

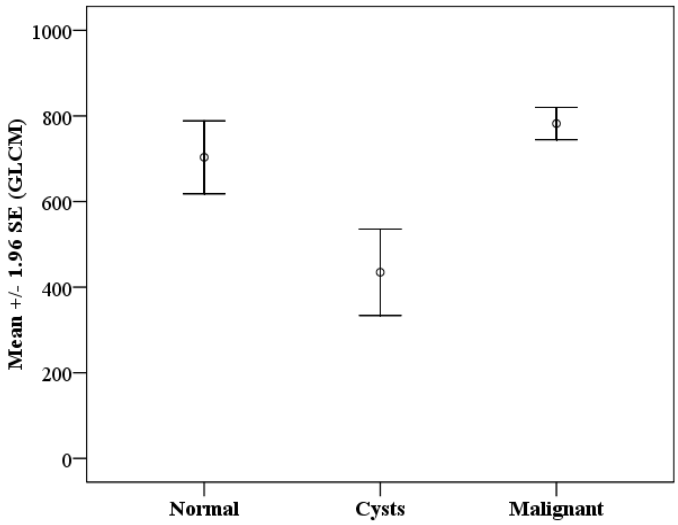


Figure 7.8: Graph showing mean values of GLCM feature for normal tissue, cysts, and malignant tissue (ROI delineation - approach 1).

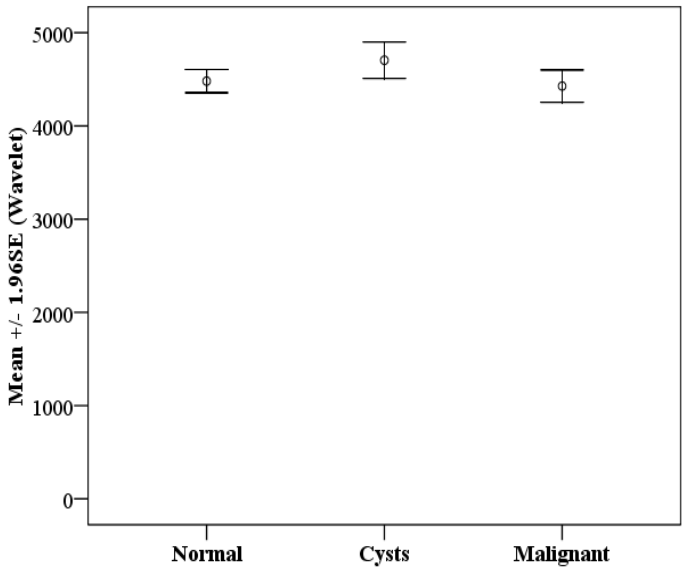


Figure 7.9: Graph showing mean values of wavelet feature for normal tissue, cysts, and malignant tissue (ROI delineation - approach 1).

Of the three groups, the malignant tissue exhibits the highest GLCM features mean value (782) while cysts exhibits the lowest GLCM mean value (435). The mean value for normal tissue is 703. On the other hand, cysts exhibit the highest wavelet features mean value (4,705) while the mean values for the normal and malignant tissues are virtually similar: 4,481 and 4,427 respectively.

The values of the extracted texture features were compared in pairs: normal tissue and cysts; normal tissue and malignant tissue; and cysts and malignant tissue. The results for the GLCM feature show that two group pairs were statistically significant: the p-value between normal tissue and cysts, and between cysts and malignant tissue, are less than 0.01. However, the mean values between normal tissue and malignant tissue were not statistically significant (p=0.12).

The results for the wavelet feature, however, show that for all three group pairs compared (normal-cysts, normal-malignant, cysts-malignant), the difference was not statistically significant: the p-value between normal tissue and cysts is 0.08, between normal tissue and malignant tissue is 0.97, and between cysts and malignant tissue is 0.07. The results for both, GLCM and wavelet features, are summarised in Table 7.2.

Table 7.2: Summary results of the significance test (ROI delineation - approach 1).

Texture Feature	**Group Pair**	**p-value**	Significantly different?
GLCM	Normal-Cysts	< 0.01	Yes
	Normal-Malignant	0.12	No
	Cysts-Malignant	<0.01	Yes
Wavelet	Normal-Cysts	0.08	No
	Normal-Malignant	0.97	No
	Cyst-Malignant	0.07	No

Receiver operating curve (ROC) analysis was performed to determine the ability of the GLCM features in discriminating between normal tissue and cyst, and between cysts and malignant tissue (i.e. for those group pairs that demonstrate a statistically significant difference (p-value < 0.05)). The corresponding ROC curves are shown in Figure 7.10 (between normal tissue and cysts) and Figure 7.11 (between cysts and malignant tissue). As mentioned in Section 3.8.4, an AUC close to 1 indicates a strong discriminatory power/ability of the indicator variable while the AUC close to 0.5 indicates that the variable has little discriminatory power.

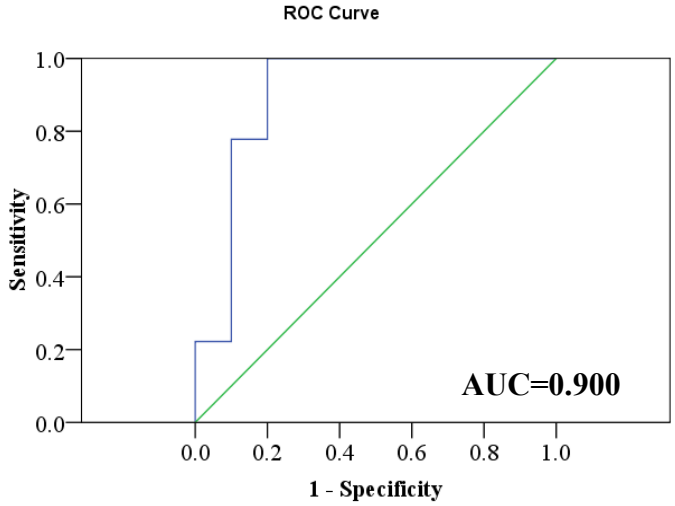


Figure 7.10: ROC curve for GLCM feature to discriminate between normal tissue and cysts (ROI delineation - approach 1).

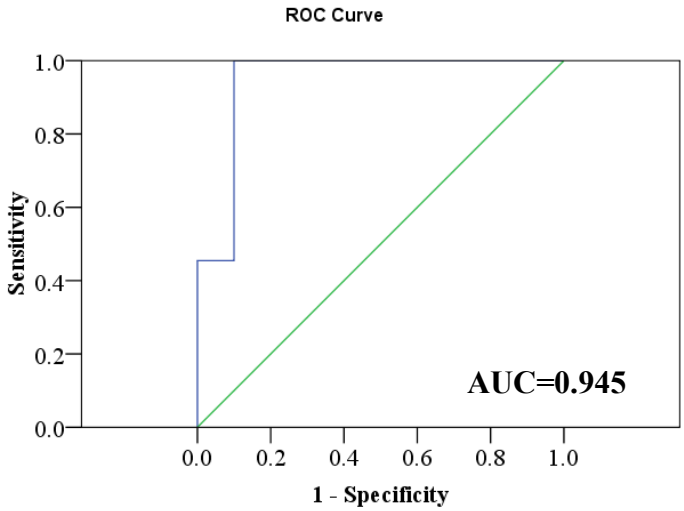


Figure 7.11: ROC curve for GLCM feature to discriminate between cysts and malignant tissue (ROI delineation - approach 1).

Table 7.3 presents the area under the curve (AUC) which reflects the ability of the GLCM feature in discriminating between the three groups, for the ROI delineation which includes the anechoic area that occurs within the normal and the malignant tissues (approach 1).

Table 7.3: The area under the curve (AUC) associated with GLCM feature (ROI delineation - approach 1).

Texture Feature	Group Pair	AUC	Discriminatory ability*
GLCM	Normal-Cysts	0.900	Good
	Cysts-Malignant	0.945	Very Good

**Based on the classification by Simon (2008)*

To evaluate the ability of the indicator variable (as in this study, the texture feature) to correctly classify images, a sensitivity of 100% and a moderately high sensitivity have been for comparison. For instance, in discriminating between normal tissue and cysts, use of 511 as a threshold value led to an estimated sensitivity of 100% and specificity of 80%. However, if the threshold value is set to 563, the sensitivity will drop to 78% while the specificity will increase to 90%. Table 7.4 presents examples of the threshold value, the sensitivity and the specificity associated with the GLCM in discriminating ovarian tissue using approach 1.

Table 7.4: Sensitivity and specificity associated with GLCM features (ROI delineation - approach 1).

Texture Features	Group Pair	Threshold Value	Sensitivity (%)	Specificity (%)
GLCM	Normal –Cysts	511	100	80
		563	78	90
	Cysts-Malignant	623	100	90
		713	82	90

7.3.2 ROI Delineation - Approach 2 (Anechoic Area Excluded)

Similar to section 7.3.1, the mean, median, standard deviation, and standard error of the mean (s.e.) of the extracted texture features were calculated for each group (normal tissue, cysts, and malignant tissue) and are summarised in Table 7.5. The mean values for the GLCM and wavelet features are visually presented in Figure 7.12 and Figure 7.13, respectively. It is worth noting that the ROI used for cyst images in this approach (approach 2) is identical to the earlier approach (approach 1); the anechoic area of the cysts is included in the analysis. As such, the mean and SD values for these two approaches are also similar.

Table 7.5: Summary of the mean, median, standard deviation, and std. error of the mean of the extracted features (ROI delineation - approach 2).

Texture Feature	Tissue Type	Mean	Median	SD	S.E.
GLCM	Normal	678	661	111	37
	Cysts	435	421	163	51
	Malignant	818	840	53	16
Wavelet	Normal	4,292	4,289	328	109
	Cysts	4,705	4,622	316	100
	Malignant	3,776	3,745	269	81

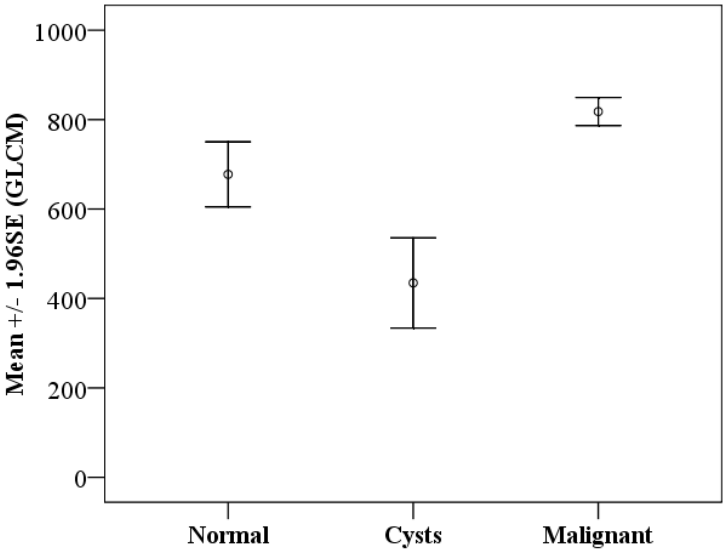


Figure 7.12: Graph showing mean values of GLCM feature for normal tissue, cysts, and malignant tissue (ROI delineation - approach 2).

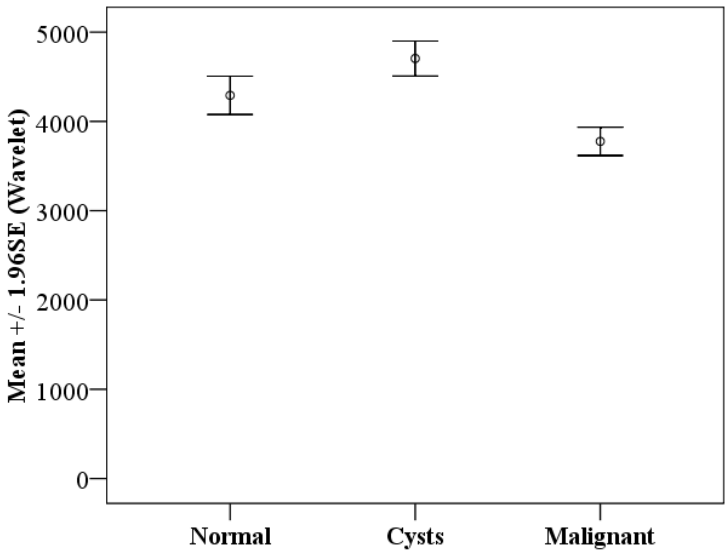


Figure 7.13: Graph showing mean values of wavelet feature for normal tissue, cysts, and malignant tissue (ROI delineation - approach 2).

Of the three groups, the malignant tissue exhibits the highest mean value for GLCM features (818) while cysts exhibits the lowest (435). The mean value for normal tissue is 678. On the other hand, regarding the wavelet feature, cysts exhibit the highest mean value (4,705) while malignant tissue exhibits the lowest mean value (3,776). The mean value for normal tissue is 4,292.

The statistical analysis results show that the mean difference for both, the GLCM and the wavelet feature, were statistically different for all three group pairs compared (normal-cyst, normal-malignant, cyst-malignant). As for the GLCM features, the p-value ranged from less than 0.001 up to 0.005, while for the wavelet feature, the p-value ranged from less than 0.001 up to 0.013. The results are summarised in Table 7.6.

Table 7.6: Summary results of the significance test (ROI delineation - approach 2)

Texture Features	Group Pair	p-value	Significantly different?
GLCM	Normal-Cysts	0.001	Yes
	Normal-Malignant	0.005	Yes
	Cysts-Malignant	<0.001	Yes
Wavelet	Normal-Cysts	0.013	Yes
	Normal-Malignant	0.002	Yes
	Cysts-Malignant	<0.001	Yes

The receiver operating curve (ROC) analysis was performed for all group pairs that demonstrate a significant difference (p-value < 0.05), to determine the ability of the GLCM and wavelet features in discriminating between the three groups. The corresponding ROC curves for the GLCM feature are shown in Figure 7.14, Figure 7.15 and Figure 7.16, while the corresponding ROC curves for the wavelet feature are shown in Figure 7.17, Figure 7.18 and Figure 7.19. The AUCs for the GLCM and wavelet features for all group pairs tested for the ROI delineation that excludes the anechoic area that occurs within the normal and the malignant tissues (approach 2) are presented in Table 7.7.

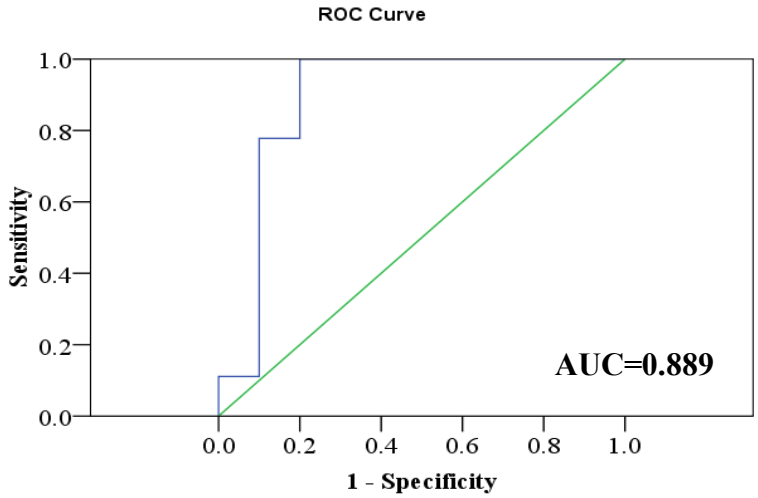


Figure 7.14: ROC curve showing the ability of GLCM feature to discriminate between normal tissue and cysts (ROI delineation - approach 2)

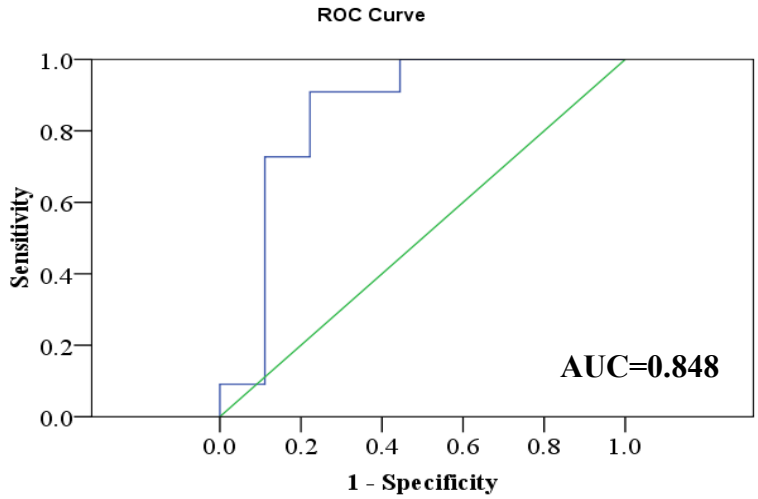


Figure 7.15: ROC curve showing the ability of GLCM feature to discriminate between normal and malignant tissues (ROI delineation - approach 2)

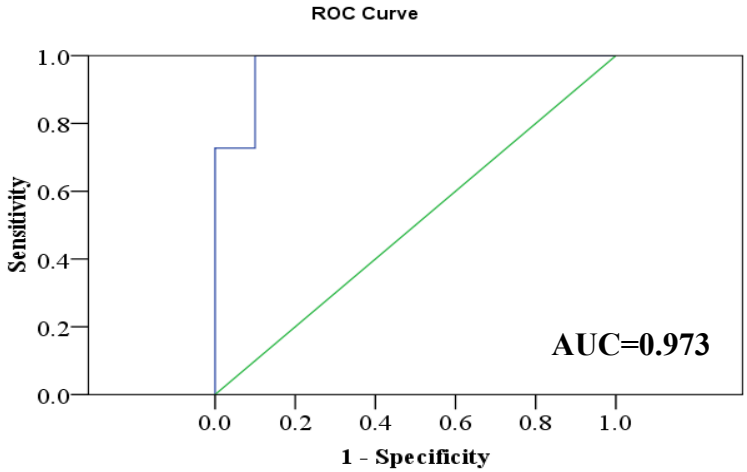


Figure 7.16: ROC curve showing the ability of GLCM feature to discriminate between cysts and malignant tissue (ROI delineation - approach 2).

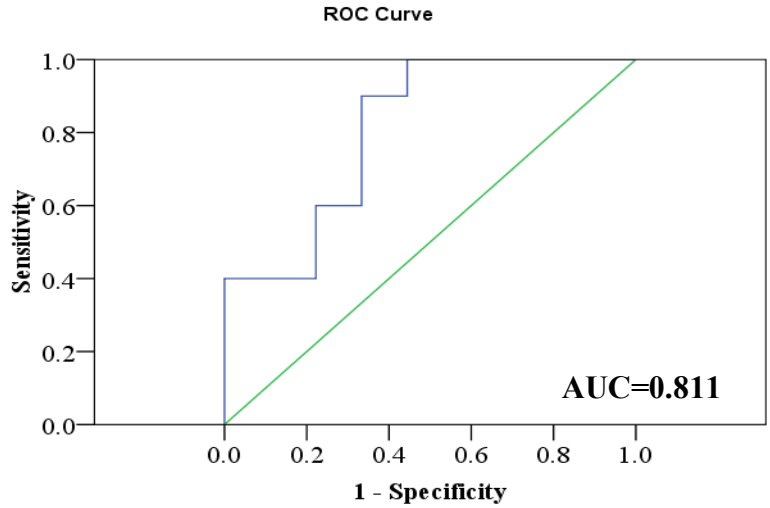


Figure 7.17: ROC curve showing the ability of wavelet feature to discriminate between normal tissue and cysts (ROI delineation - approach 2).

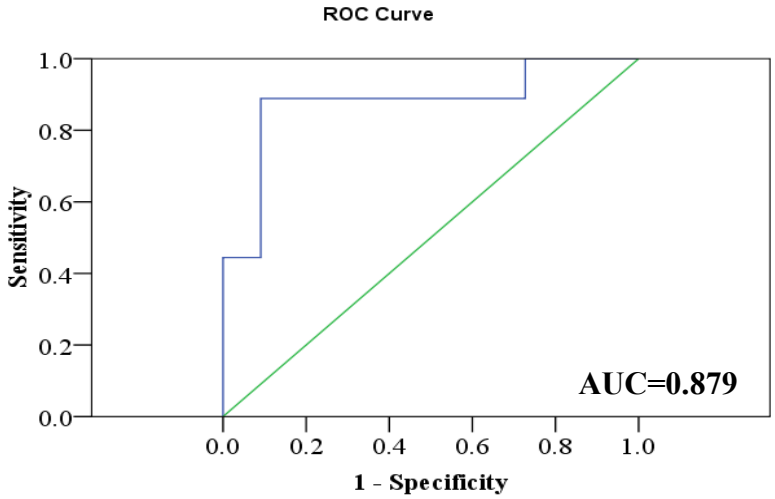


Figure 7.18: ROC curve showing the ability of wavelet feature to discriminate between normal and malignant tissues (ROI delineation - approach 2).

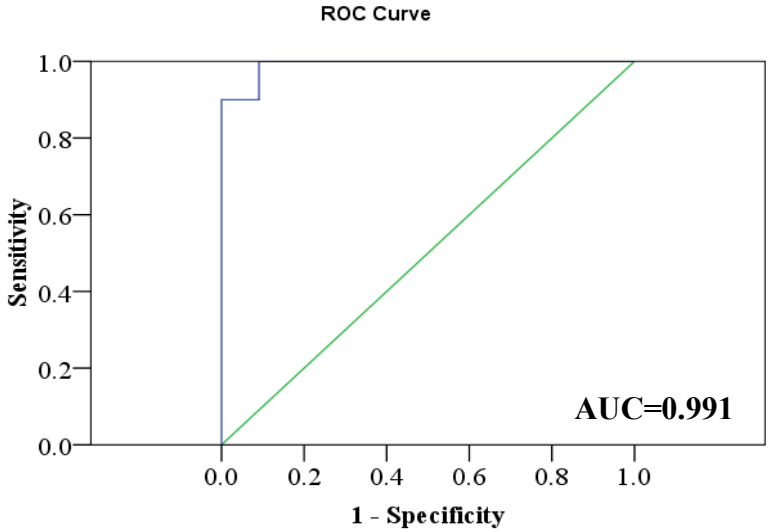


Figure 7.19: ROC curve showing the ability of wavelet feature to discriminate between cysts and malignant tissue (ROI delineation - approach 2).

Table 7.7: The area under the curve (AUC) associated with GLCM and wavelet features (ROI delineation - approach 2)

Texture Features	Group Pair	AUC	Discriminatory Ability*
GLCM	Normal-Cysts	0.889	Good
	Normal-Malignant	0.848	Good
	Cysts-Malignant	0.973	Excellent
Wavelet	Normal-Cysts	0.811	Good
	Normal-Malignant	0.879	Good
	Cysts-Malignant	0.991	Excellent

**Based on the classification by Simon (2008)*

To evaluate the ability of the GLCM and wavelet features to correctly classify images, similar to Section 7.3.1, a sensitivity of 100% and a moderately high sensitivity were chosen for comparison. For instance, in discriminating between normal and malignant tissues using the GLCM feature, use of 737 as a threshold value led to an estimated sensitivity of 91% and specificity of 78%. As for the wavelet feature, use of 4,054 as a threshold value led to an estimated sensitivity of 91% and specificity of 89%. Table 7.8 presents examples of the sensitivity and the specificity for various group pairs associated with the GLCM and wavelet features using approach 2.

Table 7.8: Sensitivity and specificity associated with GLCM and wavelet features (ROI delineation - approach 2).

Texture Features	Group Pair	Threshold Value	Sensitivity (%)	Specificity (%)
GLCM	Normal –Cysts	520	100	80
		569	78	90
	Normal-Malignant	691	100	56
		737	91	78
	Cysts-Malignant	637	100	90
		731	91	90
Wavelet	Normal –Cysts	4,313	100	56
		4,473	80	67
	Normal-Malignant	3,627	100	44
		4,054	89	91
	Cysts-Malignant	4,188	100	90
		4,362	90	90

7.4 Discussion

Ovarian cancer remains a leading cause of death from gynaecologic malignancies. The malignancy of suspicious lesions, however, has to be confirmed by a biopsy which is invasive in nature. Ultrasound imaging, on the other hand, is a non-invasive method and is widely used for ovarian cancer diagnosis. Improvements in the diagnostic accuracy of ultrasound based diagnoses will reduce the number of unnecessary biopsies.

Ultrasonic tissue characterisation describes the micro structure of a tissue either by its physical properties (such as speed and attenuation of sound) or by its echo texture (Valcx & Thijssen, 1997). It can be used in conjunction with a visual interpretation of an ultrasound image to improve the diagnosis. Nevertheless, as discussed in previous chapters, the texture analysis of ultrasound images has some disadvantages such as the dependency on the ROI size, ROI depth and gain settings. However, the statistical summary approach (i.e. mean) that we adopted in this study was able to produce texture descriptors which are robust to those possible influencing factors, as demonstrated in chapter 3, chapter 4, chapter 5, and chapter 6.

The aim of this chapter was to determine the ability of GLCM and wavelet features in characterising ovarian tissues. In particular, we investigated the appropriateness of these features in discriminating normal tissue, cysts, and malignant tissue.

There were variations observed in the texture features value extracted from the same tissue type. For example, the SD of the extracted values for GLCM feature for normal tissue was 130 for approach 1 and 111 for approach 2. Similarly, the SD for wavelet features was 191 and 328 for approach 1 and approach 2 respectively. Besides the

variations caused by the ultrasound system and the scanning process, another possible factor that may contribute to this variation is the difference in the tissue layer between different patients (Coleman, 2007), which in turn caused the level of attenuation to vary from patient to patient (Wilhjelm *et al.*, 1998; Xiao *et al.*, 2002).

In this study, two ROI selection approaches were adopted. It can be seen from the plots that omitting the anechoic area that occurs within the normal and the malignant tissues has improved the separation in the extracted texture features between different groups. This is confirmed by the significance test results whereby for approach 2, all three group pairs (normal tissue-cysts, normal tissue-malignant tissue, and cysts-malignant tissue) were significantly different for both texture features, compared to only two group pairs (normal tissue-cysts and cysts-malignant tissue) of the GLCM feature which were significantly different for approach 1, indicating that the anechoic area within the area of interest may affect the texture analysis technique.

For wavelet features, there was no statistically significant difference ($p=0.97$) between normal and malignant tissues when the whole area was included for features extraction. On the other hand, a significant difference was observed (p-value = 0.002) when the anechoic area was excluded from the normal and malignant tissues. The results indicate that the determination of the area in which texture features are extracted will affect the characterisation ability.

According to Diamond *et al.* (2004), texture analysis is more appropriate for the characterisation of regions exhibiting greater homogeneity in their structure. This suggests that for any texture analysis application meant for biological tissue

characterisation, the anechoic area should be treated with care. Failure to do so may results in false classification and affects the overall classification performance. This is particularly important when the anechoic area is very likely to be part of the structure studied, such as the ovarian follicle in the case of an ovary.

In this study, the ROC curve analysis has been used to evaluate the ability of the GLCM and wavelet features in classifying the B-mode ovarian images. The AUC, which is obtained from the ROC curve analysis, is useful parameter for evaluating the overall ability of the indicator variable (as in this study, the texture feature) to correctly classify subjects (Morrison et al., 2003). Generally, the results of this study demonstrate that the classification ability of the GLCM and wavelet features was good with the AUC ranging from 0.811 to 0.991. These findings are comparable to the results reported by van Ginneken *et al.* (2002) and Huang et al. (2008).

In their study to discriminate the radiological images for the chest screening against tuberculosis, van Ginneken *et al.* (2002) reported that the area under the ROC curve was 0.820 and concluded that the technique used may be helpful to radiologists for reading mass chest screening images. Similarly, in their study to classify breast lesions on ultrasound, Huang *et al.* (2008) reported that the area under the ROC curves was 0.91 and concluded that the proposed method provides a clinically useful second opinion.

AHRQ (2006) reported that for morphologic appearance on ultrasound, pooled sensitivities for specific scoring systems ranged from 0.82 to 0.91, and specificities ranged from 0.68 to 0.81. In discriminating between normal and malignant tissues, the results of this study show that for the GLCM feature, use of 737 as a threshold value led

to an estimated sensitivity of 91% and specificity of 78%, while for the wavelet feature, use of 4,054 as a threshold value led to an estimated sensitivity of 89% and specificity of 91%. To improve the performance of texture analysis, these two features can be applied simultaneously. According to Singh and Singh (2002), a selection of features from a combined set of textural features categories yields a better performance compared to using features from a single category, while AMD, in their document *'A New Technique (HistoScanning^{TM}) for the Non-Invasive Detection of Cancers of Solid Organs'*, reported that an improvement in the specificity was achieved when three algorithms were applied simultaneously compared with using each algorithm separately. Results from texture analysis techniques can also be combined with the results from other objective image analysis techniques (such as automated morphological features analysis) as an input to the CAD system to further improve the diagnostic accuracy.

Apart from that, Doi (2007) suggested two measures in order to improve the performance of CAD: (i) disregard obvious computer false-positive cases that can be identified by a physician easily, and (ii) incorporate the results of the computer output for subtle lesions which might be difficult for a physician to detect.

According to Materka (2004) and Tesar *et al.* (2008), the GLCM feature is commonly used in the 2D texture analysis of medical images. The results from this study demonstrated that in general, the GLCM feature has a better characterisation ability compared to the wavelet feature. This is in accordance with the statement by Tuceryan and Jain (1998): the GLCM features generally outperform other features. Likewise, Garra *et al.* (1993) in their study on breast lesions reported that GLCM is the most useful feature while Lam (1996) stated that GLCM has long been a powerful tool for

texture analysis. In terrain classification, Weszka *et al.* (1976) compared the performance of several texture features and concluded that GLCM features were the best among the texture features studied.

The performances of the texture analysis methods are determined by how they formulate the relationship between image elements (Al-Kadi, 2009). Since how a texture analysis method calculates the image texture was not the primary objective of this study, we did not investigate what the underlying factors are that cause the difference in the performance of the GLCM and wavelet features. Furthermore, according to Lerski (2006), the whole relationship of texture and its measurement to tissue structure is a very complex issue which is not fully understood.

It is important to note the difference between the automated computer diagnosis and the computer-aided diagnosis (CAD) concept. The main difference between the two concepts is the way in which the results of the computer output are utilised for the diagnosis. In an automated computer diagnosis, the diagnosis is based completely on the results of the computer output (Doi, 2007; Giger *et al.*, 2008). On the other hand, for CAD, the computer output is used as an aid to the physician in interpreting the images. Doi *et al.* (1999) defined CAD as a diagnosis made by a physician who takes into account the results of the computer output as a "second opinion", while Chen and Hsiao (2008) defined CAD as a diagnostic aid that takes into account equally the role of the physician and the benefits of the computer system.

According to Huynen *et al.* (1994), texture analysis may improve diagnostic accuracy by providing a more reproducible result and information that is difficult for humans to

perceive. In a study on breast lesions, Garra *et al.* (1993) stated that the texture analysis of ultrasound images markedly reduces the number of benign lesion biopsies without missing additional malignancies. This study demonstrated that texture analysis methods, GLCM and wavelet features in particular, are potentially able to discriminate between normal and pathologic ovarian tissue. It is worth noting that the threshold value should be obtained by performing an ROC curve analysis on a larger sample size that is represented by the groups studied before any threshold value can be set for clinical studies.

According to Doi (2007), in CAD, the computer output could be utilized by radiologists, but would not replace them. Similarly, Bader *et al.* (2000) concluded that a computer texture analysis system is able to improve the subjective assessment of ultrasound images of the breast but cannot replace it. Therefore, the results from this chapter suggest that GLCM and wavelet could potentially be used as an input for a CAD system to aid ovarian cancer diagnosis.

7.5 Summary

Many studies have demonstrated the ability of texture analysis techniques to characterise between normal and diseased tissues. In this chapter, the feasibility of the GLCM and wavelet features to characterise ovarian images has been investigated. Based on the findings from previous chapters, these two features are deemed to be robust and demonstrate good repeatability. Two ROI segmentation approaches were adopted in this study: (i) the whole area of the tissue of interest was included; and (ii) the anechoic area within the normal and malignant tissues was excluded.

Images were classified into three groups: normal ovary, cysts, and malignant tissue. The Q-Q plot was used to determine the normality of the data distribution. For the statistical analysis, the t-test and WMW test was used to compare the mean, i.e. to determine whether the mean value is different between the groups compared. The results demonstrated the feasibility of texture analysis in characterising ovarian tissue. The results also demonstrated that the anechoic area within the structure studied is likely to have an impact on the classification performance of the texture analysis techniques: the performance of the texture analysis was improved when the anechoic area within the normal and malignant tissues was excluded from features extraction. Consequently, care should be taken with the anechoic area inside the ROI in order to improve the discriminatory capability of the derived features.

Overall, this study demonstrated that there is a significant difference in the GLCM values between normal, cysts, and malignant tissues. Besides GLCM, the wavelet feature was also able to discriminate the three groups when the anechoic area was excluded in the features extraction. Ultimately, these texture features are anticipated to

be feasible to be used for a quantitative image analysis of ovarian images. Results from the texture analysis technique can be combined with a decision-making algorithm to form a CAD system. The output of this technique can serve as a second opinion for a physician, particularly for ambiguous cases, and in turn, is able to improve the overall diagnosis.

Chapter 8
Summary and Conclusions

8.1 Summary

Accurate and early diagnosis is among the crucial factors in the management of ovarian cancer as it will determine type of treatment as well as prognosis. A blood test (CA-125) is normally used in conjunction with radiological findings for the prediction of malignancy. Among the various type of imaging modalities, ultrasound, B-mode imaging in particular, is considered the main imaging procedure for the triage of ovarian cancer. Despite its proven usefulness, the main issue is that it is operator-dependent, and thus the accuracy and reproducibility of the diagnosis very much subject to the experience of the operator.

In order to reduce the operator-dependency, texture analysis, which is able to quantitatively characterise tissue through texture content, has been considered in this study. Image texture analysis has been applied in many imaging techniques, from satellite photography to document processing. In the medical domain, extensive research works have been carried out to investigate the use of the texture analysis technique in the characterisation and discrimination of biological tissue, such as liver, breast, thyroid, prostate, carotid plaques and many others, and has been proved to be valuable.

As mentioned in chapter 1, the aim of this study was to develop an objective method to be used in ultrasound-based assessment for ovarian tumours, using a texture analysis technique on a greyscale (B-mode) ultrasound image, with a view to providing a tool that is able to assist clinicians in their diagnosis of ovarian cancer.

It is important to note that it is necessary to establish the reliability of the extracted features in order to ensure the accuracy of texture analysis techniques when used for image classification. The reliability of the extracted features might be influenced by, among other factors, the ultrasound system, the operator, and the features extraction procedure.

Any instrument will have a random variation in its measurement. A series of images acquired under identical conditions will therefore vary due to these random variations in the ultrasound system. As well as the ultrasound system, the operator is another source of variation in the acquired images. It is therefore important to establish the reliability of B-mode images. Additionally, the repeatability of ROI segmentation during features extraction will also influence the reliability of the extracted features.

For these reasons, before texture features were selected for classification, ROI size, ROI depth, scanner gain setting, scanner repeatability, operator repeatability, and ROI segmentation repeatability were first considered. In addition, the influence of the 'calliper line', a line that is used to measure the size of the lesion under 'conventional' methods of interpreting B-mode images, was also assessed. These issues were addressed in Chapters 4, 5 and 6 respectively. The results are summarised in Table 8.1. Based on the results obtained, it is concluded that technical aspects of image acquisition and features extraction are adequate to allow further investigation of the ability of texture analysis technique to characterise ovarian tissue.

Table 8.1: Summary of the assessment of the B-mode image texture reliability.

	Textural Features			
Type of Assessment	**GLCM**	**RLM**	**Gradient**	
1) (a) ROI Size – Intra-ROI	✓	✓	**> 900 pixels**	
1) (b) ROI Size – Inter-ROI	✓	✗	✓	
2) (a) Depth: 4cm focus	✓	✗	✗	
2) (b) Depth: 6cm focus	✓	✗	✗	
2) (c) Depth: 8cm focus	✓	✗	✗	
3) (a) Gain: 80 to 100	✓	✗	✗	
3) (b) Gain: 90 to 100	✓	✗	✓	
4) Ultrasound system reliability	✓	✓	✓	
5) Intra-operator repeatability	✓	✓	✓	
6) Inter-operator reproducibility	✓	✓	✓	
7) ROI segmentation repeatability	✓	✓	✓	
8) Clinical image acquisition	✓	✓	✓	
9) Calliper-line	✓	✓	✗	

* ***'✓'= Not affected, '✗'= Affected***

Then, in Chapter 7, the feasibility of texture analysis technique was investigated. The texture value of three types of ovarian tissue (normal, cysts, and malignant) were compared. Two ROI segmentation approaches were adopted: (i) with the whole area of the tissue of interest included; and (ii) with the anechoic area within the area of the tissue of interest excluded.

The results revealed that the anechoic area within the structure studied is likely to have an impact on the classification performance. The results demonstrated that there was a significant difference in the GLCM and wavelet features between those groups when the anechoic area was excluded from the normal and malignant tissues. However, when the whole area of the tissue of interest was included for features extraction, there was no significant difference between normal and malignant tissue in the GLCM and wavelet features. Similarly, there was no significant difference in the wavelet feature for all the three groups compared.

It is well documented that B-mode images contain a significant level of image noise and distortions. Therefore, in this study, an image enhancement technique using Adobe Photoshop was employed to prepare the clinical images prior to the ROI segmentation and features extraction. In addition, semi-automatic segmentation (*magnetic lasso tool*) was employed to facilitate the segmentation process.

8.2 Conclusions

This study has shown that quantitative texture analysis of B-mode images using the statistical summary approach (i.e. the mean of all statistical values generated by an algorithm) demonstrates a significant difference between normal tissue, cysts, and malignant tissue for GLCM and wavelet features. Therefore, it is possible to conclude that the development of CAD using GLCM and wavelet features has the potential to differentiate objectively the ovarian lesion and this warrants further study on a larger scale of data. It is expected that this method would ultimately improve the diagnosis of medical imaging in general and ovarian cancer in particular.

This study has also demonstrated that some texture features are dependent on the ROI size, the ROI depth, and the gain setting. As for the B-mode image reliability, the phantom study results proved that ultrasound systems are able to produce consistent images and so good intra- and inter-operator reliability is achievable. Assessment of clinical image acquisition also demonstrated that a low variation in clinical image acquisition is achievable.

The results of this study also demonstrate that the use of the *magnetic lasso tool* facilitates the reproducibility of ROI segmentation procedure of B-mode images.

8.3 Contributions

A novel approach of using summary statistics of extracted texture features has been presented and successfully applied to ovarian images. According to Tuceryan and Jain (1998), one of the drawbacks of GLCM is that there is no well-established method of selecting the displacement vector (*d*), and computing co-occurrence matrices for all possible different '*d*' values is not feasible. Apart from that, according to Livens *et al.* (1997), although a large number of features may carry more information, it can also cause the classification to become more complex. They further emphasise that although there are feature reduction methods available, a fundamental problem with them is that the predominant feature that will carry the most useful information can differ from one texture to another. The approach introduced in this study is able to overcome these problems.

This study has conducted a thorough assessment on the possible influencing factors / conditions that may affect the extracted features. This included an assessment of factors associated with the ultrasound system, operator, and the feature extraction process. The results of this study can be used for similar studies in the future. Furthermore, the assessment methods applied in this study can be used to carry out similar assessments on other texture analysis methods not included in this study.

To my knowledge, there is no single study that combines both the use of the *gradient filters* (for image enhancement) and the *magnetic lasso tool* (for ROI delineation) prior to features extraction. This study demonstrated that the combination of these two techniques helps in providing good ROI segmentation reproducibility, one of the key factors that will determine the success of texture analysis in the medical domain.

8.4 Future works

Five texture features (GLCM, RLM, gradient, AR, and wavelet) were considered in this study. Out of these five features, only two (GLCM and wavelet) were found to be robust and reproducible, and deemed to be suitable for image classification. This study also demonstrated that there were significant differences between the groups compared for both GLCM and wavelet features.

However, there are some areas in which the study and the developed methods could be further improved or explored. This includes the following:

(i) A larger number of images are required in order to get more representative results and to confirm the clinical value of this method.

(ii) Ovarian cancer is classified into four stages based on the FIGO system. Future work may involve the classification of malignant tissues at different stages.

(iii) It would be of interest to explore other types of two-dimensional (2D) texture analysis methods as well as three dimensional (3D) texture analysis methods in discriminating ovarian tissue.

(iv) Wall thickness and septa are among the morphological features used in the scoring system to diagnose ovarian cancer from B-mode ultrasound images. The image enhancement techniques and *magnetic lasso tool* employed in this study could be applied to aid in the evaluation of the thickness of the wall and the septa.

References:

Abella, M. et al. 2009. Automatic quantification of histological studies in allergic asthma. *Cytometry* 75(3), pp. 271-277.

ACS. 2007. *Cancer Facts & Figures 2007*. Atlanta: American Cancer Society.

ACS. 2011. *How is Ovarian Cancer Diagnosed?* [Online]. American Cancer Society. Available at: http://www.cancer.org/Cancer/OvarianCancer/DetailedGuide/ovarian-cancer-diagnosis [Accessed: 31 October 2011].

AHRQ. 2006. *Management of Adnexal Mass*. Rockville: Agency for Healthcare Research and Quality.

Al-Kadi, O. S. 2009. *Tumour Grading and Discrimination based on Class Assignment and Quantitative Texture Analysis Techniques.* University of Sussex.

Alacam, B. et al. 2003. Breast cancer detection based on ultrasound B-scan texture analysis and patient age information. In: *Proceedings of IEEE 29th Annual Bioengineering Conference, 2003.* 22-23 March 2003. pp. 98-99.

Alcazar, J. 2006. Tumor angiogenesis assessed by three-dimensional power Doppler ultrasound in early, advanced and metastatic ovarian cancer: A preliminary study. *Ultrasound in Obstetrics and Gynecology* 28, pp. 325-329.

Alcazar, J. L. and Castillo, G. 2005. Comparison of 2-dimensional and 3-dimensional power-Doppler imaging in complex adnexal masses for the prediction of ovarian cancer. *American Journal of Obstetrics and Gynecology* 192, pp. 807-812.

Alcazar, J. L. and Jurado, M. 2011. Three-dimensional ultrasound for assessing women with gynecological cancer: A systematic review. *Gynecologic Oncology* 120(3), pp. 340-346.

Aletti, G. D. et al. 2007. Current management strategies for ovarian cancer. In: *Symposium on solid tumour.* pp. 751-770.

Alqahtani, M. 2010. *Quantitative Ultrasound Image Analysis of the Gastrocnemius Muscle for Injury Evaluation - A Pilot Study.* Cardiff University.

Alqahtani, M. et al. 2010. *Tissue characterization: Influence of ultrasound setting on texture features in vivo*. International Conference on Medical Image Analysis and Clinical Applications (MIACA), 2010 10-13 June 2010.

Altman, D. G. and Bland, J. M. 1994. Diagnostic tests 3: Receiver operating characteristic plots. *British Medical Journal* 309, p. 188.

Alvarenga, A. V. et al. 2007. Classifying breast tumours on ultrasound images using a hybrid classifier and texture features. In: *IEEE International Symposium on Intelligent Signal Processing, 2007.* 3-5 Oct. 2007. pp. 1-6.

AMD, A. M. D. 2010(a). *Prostate HistoScanning* [Online]. Advanced Medical Diagnostics. Available at: http://www.histoscanning.com/products/prostate-histoscanningtm/ [Accessed: 14 December 2010].

AMD, A. M. D. 2010(b). *Ovarian HistoScanning* [Online]. Advanced Medical Diagnostics. Available at: http://www.histoscanning.com/products/in-development/ [Accessed: 14 December 2010].

Ameye, L. et al. 2009. A scoring system to differentiate malignant from benign masses in specific ultrasound-based subgroups of adnexal tumors. *Ultrasound in Obstetrics and Gynecology* 33(1), pp. 92-101.

Averkiou, M. A. et al. 1997. A new imaging technique based on the nonlinear properties of tissues. In: *Proceedings of IEEE Ultrasonics Symposium, 1997.* 5-8 Oct 1997. pp. 1561-1566 vol.1562.

Bader, W. et al. 2000. Does texture analysis improve breast ultrasound precision? *Ultrasound in Obstetrics and Gynecology* 15(4), pp. 311-316.

Bailey, E. et al. 2007. Validation of sperm counting methods using limits of agreement. *Journal of Andrology*. p. 10.

Barnett, S. B. ed. 2000. *Ultrasound-induced Heating and Its Biological Consequences.* London: British Institute of Radiology, p. 120.

Baroni, M. et al. 2007. Towards quantitative analysis of retinal features in optical coherence tomography. *Medical Engineering & Physics* 29(4), pp. 432-441.

Bassani, D. et al. 2007. Use of the limits of agreement approach in periodontology. *Oral Health & Preventive Dentistry* 5(2), pp. 119-124.

Basset, O. et al. 1993. Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrices. *Ultrasonic Imaging* 15(3), pp. 218-237.

Benacerraf, B. R. 2008. *Three-Dimensional Ultrasound in Gynecology: Where is it Applicable?* [Online]. GE Healthcare. Available at: http://www.gehealthcare.com/usen/ultrasound/products/msucme3d_gyn.html [Accessed: 15 January 2011].

Benedet, J. et al. 2000. *Staging Classifications and Clinical Practice Guidelines for Gynaecological Cancers*. FIGO Committee on Gynecologic Oncology.

Bharati, M. H. et al. 2004. Image texture analysis: methods and comparisons. *Chemometrics and intelligent laboratory systems* 72(1), pp. 57-71.

Bland, M. J. 2000. *An Introduction to Medical Statistics*. 3rd. ed. New York: Oxford University Press, p. 405.

Bland, M. J. and Altman, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327(8476), pp. 307-310.

Bland, M. J. and Altman, D. G. 2010. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies* 47, pp. 931–936.

Board of the Faculty of Clinical Radiology. 2005. *Standards for Ultrasound Equipment*. London: The Royal College of Radiologists, p. 19.

Bocchi, L. et al. 1997. Tissue characterization from x-ray images. *Medical Engineering & Physics* 19(4), pp. 336-342.

Bonilha, L. et al. 2003. Texture analysis of hippocampal sclerosis. *Epilepsia* 44(12), pp. 1546–1550.

Boote, E. J. 2003. Doppler US techniques: Concepts of blood flow detection and flow dynamics. *RadioGraphics* 23(5), pp. 1315-1327.

Booth, B. et al. 2006. Towards medical ultrasound image segmentation with limited prior knowledge. In: *IEEE 12th - Signal Processing Education Workshop.* 24-27 Sept. 2006. pp. 488-493.

Boukerroui, D. et al. 1998. Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation
European Journal of Ultrasound
8(2), pp. 135-144.

Braeckman, J. et al. 2007. Computer-aided ultrasonography (HistoScanning): A novel technology for locating and characterizing prostate cancer. *BJU International* 101(3), pp. 293-298.

Braeckman, J. et al. 2008. The accuracy of transrectal ultrasonography supplemented with computer-aided ultrasonography for detecting small prostate cancers. *BJU International* 102(11), pp. 1560-1565.

Bridal, S. L. et al. 2003. Milestones on the road to higher resolution, quantitative, and functional ultrasonic imaging. In: *Proceedings of the IEEE Vol. 91, No. 10, October 2003.* pp. 1543-1561.

Browne, J. E. et al. 2003. *Assessment of the acoustic properties of common tissue-mimicking test phantoms* [Online]. Available at: http://arrow.dit.ie/scschphyart/10 [Accessed. 15 January 2011]

Browne, J. E. et al. 2004. Objective measurements of image quality. *Ultrasound in Medicine & Biology* 30(2), pp. 229–237.

Bruton, A. et al. 2000. Reliability: What is it, and how is it measured? *Physiotherapy* 86(2), pp. 94-99.

Cancer Research UK. 2009(a). *Ovarian cancer* [Online]. Available at: http://info.cancerresearchuk.org/utilities/atozindex/atoz-ovarian-cancer [Accessed: 28 October 2010].

Cancer Research UK. 2009(b). *Ovarian cancer symptoms, diagnosis and treatment* [Online]. Available at: http://info.cancerresearchuk.org/cancerstats/types/ovary/symptoms/#symptoms [Accessed: 28 October 2010].

Cancer Research UK. 2010(a). *Key Facts: Ovarian Cancer* [Online]. Available at: http://info.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@sta/documents/generalcontent/crukmig_1000ast-3058.pdf [Accessed: 06 January 2011].

Cancer Research UK. 2010(b). *Ovarian cancer risk factors* [Online]. Available at: http://info.cancerresearchuk.org/cancerstats/types/ovary/riskfactors/ [Accessed: 06 January 2011].

Cancer Research UK. 2010(c). *Ovarian cancer survival statistics* [Online]. Available at: http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/ [Accessed: 06 January 2011].

Cancer Research UK. 2010(d). *Stages of ovarian cancer* [Online]. Available at: http://www.cancerhelp.org.uk/type/ovarian-cancer/treatment/stages-of-ovarian-cancer [Accessed: 15 January 2011].

Castellano, G. et al. 2004. Texture analysis of medical images. *Clinical Radiology* 59(12), pp. 1061-1069.

Catherine, S. et al. 2006. Quantitative image analysis in sonograms of the thyroid gland. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 569(2), pp. 606-609.

Chambless, L. E. et al. 1996. Variability in B-mode ultrasound measurements in the Atherosclerosis Risk in Communities (ARIC) study. *Ultrasound in Medicine & Biology* 22(5), pp. 545-554.

Chan, H. P. et al. 1995. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. *Physics in Medicine and Biology* 40(5), pp. 857-876.

Chan, K. K. and Selman, T. J. 2006. Testing for ovarian cancer. *Best Practice & Research Clinical Obstetrics & Gynaecology* 20(6), pp. 977-983.

Chan, K. L. 1995. Quantitative characterization of carotid atheromatous plaques using ultrasonic RF A-scan
data. *Ultrasonics* 33(3), pp. 163-164.

Chan, K. L. 1998. Adaptation of ultrasound image texture characterization parameters. In: *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE.* 29 Oct-1 Nov 1998. pp. 804-807 vol.802.

Chan, K. L. and McCarty, K. 1990. Aspects of the statistical texture analysis of medical ultrasound images. In: *IEE Colloquium on Ultrasound Instrumentation.* 10 May 1990. pp. 1/3-3/3.

Chen, D.-R. and Hsiao, Y.-H. 2008. Computer-aided Diagnosis in Breast Ultrasound. *Journal of Medical Ultrasound* 16(1), pp. 46-56.

Chen, G. et al. 2005. Evaluation of antivascular therapy with texture analysis. *Anticancer Research* 25(5), pp. 3399-3405.

Chen, W. et al. 2010. Texture Analysis of Brain CT Scans for ICP Prediction. In: Elmoataz, A. et al. eds. *Image and Signal Processing*. Vol. 6134. Springer Berlin / Heidelberg, pp. 568-575.

Cheng, H. D. et al. 2010. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition* 43(1), pp. 299-317.

Chien, Y. P. and Fu, K.-S. 1974. Recognition of x-ray picture patterns *IEEE Transactions On Systems, Man, And Cybernetics* SMC-4(2), pp. 145-156.

Christodoulou, C. I. et al. 2003. Texture-based classification of atherosclerotic carotid plaques. *IEEE Transactions on Medical Imaging* 22(7), pp. 902-912.

Claudon, M. et al. 2002. Advances in ultrasound. *European radiology* 12(1), pp. 7-18.

Coleman, D. 2007. *Characterisation of carotid atherosclerotic plaque composition in vivi using 3D B-mode ultrasound.* Cardiff Uniersity.

Coleman, D. P. et al. 2005. Development and validation of an in vivo analysis tool to identify changes in carotid plaque tissue types in serial 3-D ultrasound scans. *Ultrasound in Medicine & Biology* 31(3), pp. 329-335.

Collewet, G. et al. 2004. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic Resonance Imaging* 22(1), pp. 81-91.

Collinger, J. L. et al. 2009. Reliability of Quantitative Ultrasound Measures of the Biceps and Supraspinatus Tendons. *Academic Radiology* 16(11), pp. 1424-1432.

Conners, R. W. and Harlow, C. A. 1980. A theoretical comparison of texture algorithms. *IEEE Transaction on Pattern Analysis and Machine Intelligence* PAMI-2(3).

Crum, C. P. ed. 2004. *The Female Genital Tract.* Seventh ed. Philadelpia, Pennsylvania: Elsevier Inc., p. 1525.

Dai, S.-Y. et al. 2008. Does three-dimensional power Doppler ultrasound improve the diagnostic accuracy for the prediction of adnexal malignancy? *Journal of Obstetrics and Gynaecology Research* 34(3), pp. 364-370.

Das, A. et al. 2008. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. *Gastrointestinal Endoscopy* 67(6), pp. 861-867.

Davis, L. S. 1980. *Image Texture Analysis Techniques - A Survey*. Texas: Department of Computer Sciences, University of Texas at Austin.

de Vet, H. C. W. et al. 2006. When to use agreement versus reliability measures. *Journal of Clinical Epidemiology* 59(10), pp. 1033-1039.

DePriest, P. et al. 1993. A morphology index based on sonographic findings in ovarian cancer. *Gynecologic Oncology* 51(1), pp. 7-11.

DePriest, P. D. et al. 1994. The efficacy of a sonographic morphological index in identifying ovarian cancer: A multi-institutional investigation. *Gynecologic oncology* 55, pp. 174-178.

Derchi, L. et al. 2001. Ultrasound in gynecology. *European Radiology* 11(11), pp. 2137-2155.

Dhawan, A. P. 2003. *Medical Image Analysis*. Hoboken,New Jersey: John Wiley & Sons, p. 315.

Diamond, J. et al. 2004. The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. *Human Pathology* 35(9), pp. 1121-1131.

Doi, K. 2007. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics* 31(4-5), pp. 198-211.

Doi, K. et al. 1999. Computer-aided diagnosis in radiology: potential and pitfalls. *European Journal of Radiology* 31(2), pp. 97-109.

Dougherty, D. 2009. *Digital Image Processing for Medical Applications*. Cambridge: Cambridge University Press.

Drukker, K. et al. 2003. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. *Medical Physics* 30, pp. 1833-1842.

Duck, F. 1990. *Physical properties of Tissue - A Comprehensive Reference Book*. London: Academic Press, p. 346.

Duck, F. A. ed. 2000. *The Propagation of Ultrasound Through Tissue.* London: British Institute of Radiology, p. 120.

Dudely, N. ed. 2010. *B-mode Measurement.* 2nd. ed. Cambridge: Cambridge University Press, p. 263.

Elatrozy, T. et al. 1998. The ojective characterisation of ultrasonic carotid plaque features. *Eur J Vasc Endovasc Surg* 16, p. 8.

Fellingham, L. L. and Sommer, F. G. 1984. Ultrasonic characterization of tissue structure in the in vivo human liver and spleen. *IEEE Transactions on Sonics and Ultrasonics* 31(4), pp. 418-428.

Fishman, D. A. et al. 2001. The role of ultrasound in detecting early ovarian carcinoma:The National Ovarian Cancer Early Detection Program *Medica Mundi* 45/2, pp. 42-47.

Flam, F. et al. 1988. Symptomatology of ovarian cancer. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 27(1), pp. 53-57.

Forstner, R. et al. 1995. Ovarian cancer: Staging with CT and MR imaging. *Radiology* 197, pp. 619-626.

Fry, T. 2002. *Evaluation of the Repeatability of Residual Stress Measurements Using X-Ray Diffraction*. Middlesex: National Physical Laboratory.

Galloway, M. M. 1975. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing* 4(2), p. 6.

Garra, B. S. et al. 1993. Improving the distinction between benign and malignant breast lesions: The value of sonographic texture analysis. *Ultrasonic Imaging* 15(4), pp. 267-285.

Giger, M. L. et al. 2008. Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. *Medical Physics* 35(12), pp. 5799–5820.

Gletsos, M. et al. 2003. A computer-aided diagnostic system to characterize CT focal liver lesions: Design and optimization of a neural network classifier. *IEEE Transactions on Information Technology in Biomedicine* 7(3), pp. 153-162.

Goff, B. A. et al. 2000. Ovarian carcinoma diagnosis: Results of a national ovarian cancer survey. *Cancer Research* 89(10), pp. 2068–2075.

Gonzalez, R. C. and Woods, R. E. 1992. *Instructor's Manual for Digital Image Processing*. Reading MA: Addison-Wesley, p. 144.

Gossage, K. W. et al. 2003. Texture analysis of optical coherence tomography images: feasibility for tissue classification. *Journal of Biomedical Optics* 8(3), pp. 570-575.

Gramellini, D. et al. 2008. Diagnostic accuracy of IOTA ultrasound morphology in the hands of less experienced sonographers. *Australian and New Zealand Journal of Obstetrics and Gynaecology* 48, pp. 195-201.

Gruszauskas, N. P. et al. 2008. Performance of breast ultrasound computer-aided diagnosis: Dependence on image selection. *Acad Radiol.* 15(10), pp. 1234-1245.

Guggenbuhl, P. et al. 2008. Reproducibility of CT-based bone texture parameters of cancellous calf bone samples: Influence of slice thickness. *European Journal of Radiology* 67(3), pp. 514-520.

Guggenbuhl, P. et al. 2006. Texture analysis of X-ray radiographs of iliac bone is correlated with bone micro-CT. *Osteoporosis International* 17(3), pp. 447-454.

Hajek, M. et al. eds. 2006. *Texture Analysis for Magnetic Resonance Imaging.* Prague: Med4publishing.

Halliwell, M. 2010. A tutorial of ultrasound physics and imaging techniques. *Journal of Engineering In Medicine* 224(H2), pp. 127-142.

Han, S. M. et al. 2008. Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image. *Journal of Digital Imaging* 21(Suppl 1), pp. S121-S133.

Hangiandreou, N. J. 2003. B-mode US: Basic concepts and new technology. *RadioGraphics* 23, pp. 1019-1033.

Haralick, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67(5), pp. 786-804.

Haralick, R. M. et al. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3(6), pp. 610-621.

Harrison, L. et al. 2008. Texture analysis on MRI images of non-Hodgkin lymphoma. *Computers in Biology and Medicine* 38(4), pp. 519-524.

Harrison, L. C. V. et al. 2010. MRI Texture analysis in multiple sclerosis: Toward a clinical analysis protocol. *Academic Radiology* 17(6), pp. 696-707.

He, Z. et al. 2004. Dependence of tissue characterization features on region of interest (ROI) size: studies on phantoms and simulations. In: *IEEE Ultrasonics Symposium, 2004.* pp. 2082-2085 Vol.2083.

Herlidou, S. et al. 2003. MRI texture analysis on texture test objects, normal brain and intracranial tumors. *Magnetic Resonance Imaging* 21(9), pp. 989-993.

Herlidou, S. et al. 1999. Comparison of automated and visual texture analysis in MRI: Characterization of normal and diseased skeletal muscle. *Magnetic Resonance Imaging* 17(9), pp. 1393-1397.

Herzog, K. et al. 2008. Examination of cyclic changes in bovine luteal echotexture using computer-assisted statistical pattern recognition techniques. *Animal Reproduction Science* 106(3-4), pp. 289-297.

Holli, K. et al. 2010. Texture analysis of MR images of patients with Mild Traumatic Brain Injury. *BMC Medical Imaging* 10:8.

Hollingsworth, K. G. and Lomas, D. J. 2005. Liver texture analysis: robustness of measurement in cirrhotic patients and healthy volunteers. In: *Proceedings International Society for Magnetic Resonance in Medicine.* p. 332.

Horskin, P. ed. 2010. *Principles of Doppler Ultrasound.* 2nd. ed. Cambridge: Cambridge University Press, p. 263.

Huang, Y. L. et al. 2008. Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. *Ultrasound in Obstetrics and Gynecology* 32(4), pp. 565-572.

Huber, S. et al. 2000. Relevance of sonographic B-mode criteria and computer-aided ultrasonic tissue characterization in differential/diagnosis of solid breast masses. *Ultrasound in Medicine & Biology* 26(8), pp. 1243-1252.

Huynen, A. L. et al. 1994. Analysis of ultrasonographic prostate images for the detection of prostatic carcinoma: The Automated Urologic Diagnostic Expert system. *Ultrasound in Medicine & Biology* 20(1), pp. 1-10.

Hwang, Y. I. et al. 2005. Techniques on semiautomatic segmentation using the Adobe Photoshop. In: *Seventh International Workshop on enterprise networking and Computing in Healthcare Industry, 2005.* 23-25 June 2005. pp. 411-414.

IEEE. 1990. *IEEE Standard Glossary of Image Processing and Pattern Recognition Terminology*. IEEE, p. 15.

Jaqaman, K. and Danuser, G. 2009. *Computational Image Analysis of Cellular Dynamics: A Case Study Based on Particle Tracking*. New York: Cold Spring Harbour.

Jeong, Y. Y. et al. 2000. Imaging evaluation of ovarian masses. *Radiographics* 20(5), pp. 1445-1470.

Jiang, J. et al. 2010. Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics* 34(8), pp. 617-631.

Jirak, D. et al. 2002. Texture analysis of human liver. *Journal of Magnetic Resonance Imaging* 15(1), pp. 68-74.

Jokubkiene, L. et al. 2007. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? *Ultrasound in Obstetrics and Gynecology* 29(2), pp. 215-225.

Jones, M. 2009. *Basics Statistics*. Cardiff: School of Mathematics, Cardiff University.

Kadah, Y. M. et al. 1996. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Transactions on Medical Imaging* 15(4), pp. 466-478.

Kakar, M. and Olsen, D. R. 2009. Automatic segmentation and recognition of lungs and lesion from CT scans of thorax. *Computerized Medical Imaging and Graphics* 33(1), pp. 72-82.

Kakkos, S. K. et al. 2007. Texture analysis of ultrasonic images of symptomatic carotid plaques can identify those plaques associated with ipsilateral embolic brain infarction. *European Journal of Vascular and Endovascular Surgery* 33(4), pp. 422-429.

Karahaliou, A. et al. 2007. Can texture of tissue surrounding microcalcifications in mammography be used for breast cancer diagnosis? *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 580(2), pp. 1071-1074.

Karanicolas, P. J. et al. 2009. Evaluating agreement: Conducting a reliability study. *The Journal of Bone and Joint Surgery* 91, pp. 99-106.

Kinkel, K. et al. 2000. US characterization of ovarian masses: A Meta-Analysis. *Radiology* 217, pp. 803-811.

Kirkwood, B. R. and Sterne, J. A. C. 2003. *Essential Medical Statistics*. Oxford: Blackwell Publishing Ltd., p. 501.

Kociołek, M. et al. 2001. Discrete wavelet transform – derived features for digital image texture analysis. In: *International Conference on Signals and Electronic Systems.* Lodz, Poland, 18-21 September 2001. pp. 163-168.

Kumar, S. S. and Moni, R. S. 2010. Diagnosis of liver tumor from CT images using fast discrete curvelet transform. *International Journal of Computer Applications*, pp. 1-6.

Kupesic, S. and Plavsic, B. M. 2006. Early ovarian cancer: 3-D power Doppler. *Abdominal Imaging* 31(5), pp. 613-619.

Kurani, A. S. et al. 2004. Co-occurrence matrices for volumetric data. In: *Computer Graphics and Imaging* Kauai, Hawaii, USA, August 16 - 18, 2004.

Kurjak, A. and Kupesic, S. eds. 2003. *Color Doppler and 3D Utrasound in Gynecology, Infertility and Obstetrics.* New Delhi: Jaypee Brothers Publishers, p. 252.

Kurjak, A. et al. 2002. Ultrasonic assessment of the peri- and postmenopausal ovary. *Maturitas* 41(4), pp. 245-254.

Kurjak, A. et al. 2003. The detection of stage I ovarian cancer by three-dimensional sonography and power Doppler. *Gynecologic Oncology* 90, pp. 258–264.

Kurtz, A. B. et al. 1999. Diagnosis and staging of ovarian cancer: Comparative values of doppler and conventional US, CT, and MR Imaging correlated with surgery and histopathologic analysis - Report of the Radiology Diagnostic Oncology Group. *Radiology* 212(1), pp. 19-27.

Laban, M. et al. 2007. Three-dimensional ultrasound and three-dimensional power Doppler in the assessment of ovarian tumors. *International Journal of Gynecology & Obstetrics* 99(3), pp. 201-205.

Lachin, J. M. 2004. The role of measurement reliability in clinical trials. *Clinical Trials* 1(6), pp. 553 - 566.

Lagalla, R. and Midiri, M. 1998. Image quality control in breast ultrasound. *European Journal of Radiology* 27(Supplement 2), pp. S229-S233.

Lam, S. W. C. 1996. Texture feature extraction using gray level gradient based co-occurence matrices. In: *IEEE International Conference on Systems, Man, and Cybernetics, 1996.* 14-17 Oct 1996. pp. 267-271 vol.261.

Laugsand, E. A. et al. 2010. Health care providers underestimate symptom intensities of cancer patients: A multicenter European study. *Health and Quality of Life Outcomes* 8(104), pp. 1-13.

Layer, G. et al. 1990. Computerized ultrasound B-scan texture analysis of experimental fatty liver disease: influence of total lipid content and fat deposit distribution. *Ultrasonic Imaging* 12(3), pp. 171-188.

Lee, W.-L. et al. 2003. Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform. *IEEE Transactions on Medical Imaging* 22(3), pp. 382-392.

Leeds Metropolitan University. 2009. *SPSS 14: Quick guide* [Online]. Leeds: Leeds Metropolitan University. Available at: http://libraryonline.leedsmet.ac.uk/lco/publications/pdf/it/qg-24.pdf [Accessed: 18 January 2011].

Lefebvre, F. et al. 2000. Computerized ultrasound B-scan characterization of breast nodules. *Ultrasound in Medicine & Biology* 26(9), pp. 1421-1428.

Lefebvre, F. et al. 1998. Texture analysis of echo-endoscopic images of the pancreas. *European Journal of Ultrasound* 7(Supplement 1), pp. S36-S36.

Lerner, J. P. et al. 1994. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. *American Journal of Obstetrics and Gynecology* 170(1 I), pp. 81-85.

Lerouxel, E. et al. 2004. Mandibular bone loss in an animal model of male osteoporosis (orchidectomized rat): a radiographic and densitometric study. *Osteoporosis International* 15(10), pp. 814-819.

Lerski, R. ed. 2006. *Clinical Application of Texture Analysis.* Prague: Med4Publishing p. 234.

Lerski, R. A. et al. 1979. Computer analysis of ultrasonic signals in diffuse liver disease. *Ultrasound in Medicine & Biology* 5(4), pp. 341-343.

Lerski, R. A. et al. 1981. Discriminant analysis of ultrasonic texture data in diffuse alcoholic liver disease: 1. Fatty liver and cirrhosis. *Ultrasonic Imaging* 3(2), pp. 164-172.

Li, B. and Meng, M. Q. H. 2009. Texture analysis for ulcer detection in capsule endoscopy images. *Image and Vision Computing* 27(9), pp. 1336-1342.

Li, H. et al. 2004(b). Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. *International Congress Series* 1268, pp. 878-881.

Li, P. S. et al. 2004(a). The reproducibility and short-term and long-term repeatability of sonographic measurement of splenic length. *Ultrasound in Medicine & Biology* 30(7), pp. 861-866.

Liu, B. et al. 2010. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. *Pattern Recognition* 43(1), pp. 280-298.

Livens, S. et al. 1997. Wavelets for texture analysis, an overview. In: *Sixth International Conference on Image Processing and Its Applications, 1997.* 14-17 Jul 1997. pp. 581-585 vol.582.

Llobet, R. et al. 2007. Computer-aided detection of prostate cancer. *International Journal of Medical Informatics* 76(7), pp. 547-556.

Loren, D. E. et al. 2002. Computer-assisted analysis of lymph nodes detected by EUS in patients with esophageal carcinoma. *Gastrointestinal Endoscopy* 56(5), pp. 742-746.

Lucidarme, O. et al. 2010. A new computer-aided diagnostic tool for non-invasive characterisation of malignant ovarian masses: results of a multicentre validation study. *European Radiology* 20(8), pp. 1822-1830.

Lugt, A. v. d. et al. 1998. Reliability and reproducibility of automated contour analysis in intravascular ultrasound images of femoropopliteal arteries. *Ultrasound in Medicine & Biology* 24(1), pp. 43-50.

Luo, L.-Y. et al. 2003. The serum concentration of human kallikrein 10 represents a novel biomarker for ovarian cancer diagnosis and prognosis. *Cancer Research* 63, pp. 807-811.

Marshall, C. 2008. *Diseases and Disorders* New York: Marshall Cavendish Corporation, p. 270.

Martin, K. ed. 2010. *Introduction to B-Mode Imaging.* 2nd. ed. Cambridge: Cambridge University Press, p. 263.

Martin, K. and Ramnarine, K. eds. 2010. *Physics.* 2nd. ed. Cambridge: Cambridge University Press, p. 263.

Mass. 1994. *Statistical Methods in Ultrasonic Tissue Characterisation.* Massachusetts Institute of Technology.

Materka, A. 2004. Texture analysis methodologies for magnetic resonance imaging. *Dialogues in Clinical Neuroscience* 6(2), pp. 243-250.

Materka, A. ed. 2006(a). *What is the texture?* Prague: Med4 Publising, p. 234.

Materka, A. 2006(b). *MaZda User's Manual*. Institute of Electronics, Technical University of Lodz.

Materka, A. et al. 1999. Texture analysis of x-ray images for detection of changes in bone mass and structure. In: *Infotech Oulu Workshop on Texture Analysis in Machine Vision.* Oulu, Finland. pp. 29-34.

Materka, A. and Strzelecki, M. 1998. *Texture Analysis Methods – A Review*. Brussels: Technical University of Lodz, Institute of Electronics.

Mathias, J. M. et al. 1999. Texture analysis of spinal cord pathology in multiple sclerosis. *Magnetic Resonance in Medicine* 42, pp. 929–935.

Mayerhoefer, M. et al. 2010. Feasibility of texture analysis for the assessment of biochemical changes in meniscal tissue on T1 maps calculated from delayed gadolinium-enhanced magnetic resonance imaging of cartilage data: comparison with conventional relaxation time measurements *Investigative Radiology* 45(9), pp. 543-547.

Mayerhoefer, M. E. et al. 2008. Are signal intensity and homogeneity useful parameters for distinguishing between benign and malignant soft tissue masses on MR images? Objective evaluation by means of texture analysis. *Magnetic Resonance Imaging* 26, pp. 1316-1322.

McDonald, J. H. 2009. *Handbook of Biological Statistics (2nd ed.)*. 2nd. ed. Baltimore, Maryland: Sparky House Publishing.

Menke, M. N. et al. 2009. Reproducibility of Retinal Thickness Measurements in Healthy Subjects Using Spectralis Optical Coherence Tomography. *American Journal of Ophthalmology* 147(3), pp. 467-472.

Merz, E. 2007. *Ultrasound in obstetrics and gynecology: Gynecology*. 2nd. ed. New York: Thieme, p. 209.

Mescam, M. et al. 2007. *Coupling texture analysis and physiological modeling for liver dynamic MRI interpretation*. Engineering in Medicine and Biology Society, 2007. 29th Annual International Conference of the IEEE. 22-26 Aug. 2007.

Michail, G. et al. 2007. Texture analysis of perimenopausal and post-menopausal endometrial tissue in grayscale transvaginal ultrasonography. *The British Journal of Radiology* 80, pp. 609-616.

Miller, P. and Astley, S. 1992. Classification of breast tissue by texture analysis. *Image and Vision Computing* 10(5), pp. 277-282.

Mojsilovic, A. et al. 1998. Characterization of visually similar diffuse diseases from B-Scan liver images using nonseparable wavelet transform *IEEE Transactions on Medical Imaging* 17(4).

Morris, D. T. 1988(a). The correction of ultrasonic image analysis features for their depth defendence. *International Journal of Bio-Medical Computing* 22(2), pp. 143-152.

Morris, D. T. 1988(b). An evaluation of the use of texture measurements for the tissue characterisation of ultrasonic images of in vivo human placentae. *Ultrasound in Medicine & Biology* 14(5), pp. 387-395.

Morrison, A. M. 2005. Receiver Operating Characteristic (ROC) Curve Preparation - A Tutorial. In: Department, E.Q. ed. Boston.

Morrison, A. M. et al. 2003. Receiver operating characteristic curve analysis of beach water quality indicator variables. *Applied and environmental microbiology* 69(11), pp. 6405-6411.

Moss, E. L. et al. 2005. The role of CA125 in clinical practice. *Journal of Clinical Pathology* 58(3), pp. 308-312.

Munzenmayer, C. et al. 2009. Texture-based computer-assisted diagnosis for fiberscopic images. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* 3-6 Sept. 2009. pp. 3735-3738.

Murmis, V. G. et al. 1988. Texture analysis of ultrasound B-scans to aid diagnosis of cancerous lesions in the breast. In: *Proceedings of IEEE Ultrasonics Symposium, 1988.* 2-5 Oct 1988. pp. 839-842 vol.832.

Myles, P. S. and Cui, J. 2007. Using the Bland–Altman method to measure agreement with repeated measures. *British Journals of Anaesthesia* 99(3), pp. 309-311.

Nailon, W. H. ed. 2010. *Texture Analysis Methods for Medical Image Characterisation.* Vukovar: InTech p. 100.

Nailon, W. H. et al. 1996. Comparative study of textural analysis techniques to characterise tissue from intravascular ultrasound. In: *International Conference on Image Processing.* 16-19 Sep 1996. pp. 303-306.

National Cancer Institute. 2006. *What You Need to Know About Ovarian Cancer*. National Institutes of Health.

National Physical Laboratory. 2005. *Measurement Good Practice Guide No. 52*. Teddington, Middlesex: National Physical Laboratory.

Nelson, T. R. 2006. Three-dimensional ultrasound imaging. In: *UIA Annual Meeting.* Ultrasonic Industry Association pp. 1-5.

Nicholas, D. et al. 1986. Tissue characterization from ultrasound B-scan data. *Ultrasound in Medicine & Biology* 12(2), pp. 135-143.

Nielsen, P. K. et al. 2000. Quantitative ultrasound image analysis of the supraspinatus muscle. *Clinical Biomechanics* 15(Supplement 1), pp. S13-S16.

Nir, D. et al. *A New Technique for The Non-Invasive Detection of Cancers of Solid Organs* [Online]. Available at: http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.131.8426 [Accessed: 10-Nov-2009].

Noble, J. A. 2010. Ultrasound image segmentation and tissue characterization. *Journal of Engineering in Medicine* 224(H2), pp. 307-316.

Noble, J. A. and Boukerroui, D. 2006. Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging* 25(8), pp. 987-1010.

Norton, I. D. et al. 2001. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. *Gastrointestinal Endoscopy* 54(5), pp. 625-629.

Olabarriaga, S. D. 2001. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis 5 (2001)* pp. 127–142.

Olson, S. H. et al. 2001. Symptoms of ovarian cancer. *Obstetrics & Gynecology* 98(2), pp. 212-217.

Oosterveld, B. J. et al. 1991. Ultrasound attenuation and texture analysis of diffuse liver disease: Methods and preliminary results. *Physics in Medicine and Biology* 36(8), pp. 1039-1064.

Ortashi. 2008. Cardiff Malignancy Index (CMI) – A novel scoring system for assessing ovarian masses. In: *BSGI Inagural Meeting 2008.* 7-9 February 2008.

Park, J. S. et al. 2005. Technical report on semiautomatic segmentation using the Adobe Photoshop. *Journal of Digital Imaging* 18(4), pp. 333-343.

Partio, M. et al. 2002. Rock texture retrieval using gray level co-occurrence matrix. In: *5th Nordic Signal Processing Symposium.* October 2002.

Pierson, R. A. and G.P.Adam. 1995. Computer-assisted image analysis, diagnostic ultrasonography and ovulation induction: Strange bedfellows. *Theriogenology* 43, pp. 105-112.

Prager, R. W. et al. 2010. Three-dimensional ultrasound imaging. *Journal of Engineering in Medicine* 224(H2), pp. 193-223.

Rakebrandt, F. et al. 2000. Relationship between ultrasound texture classification images and histology of atherosclerotic plaque. *Ultrasound in Medicine & Biology* 26(9), pp. 1393-1402.

Ranggayan, R. M. 2005. *Biomedical Image Analysis*. New York: CRC Press, p. 1271.

Ribeiro, R. and Sanches, J. eds. 2009. *Fatty Liver Characterization and Classification by Ultrasound.* Springer Berlin / Heidelberg, pp. 354-361.

Rieck, G. C. et al. 2006. Power Doppler in the assessment of pelvic masses in a low risk group. *Journal of Obstetrics and Gynaecology* 26(3), pp. 222-224.

Rocha, R. et al. 2011. Segmentation of ultrasound images of the carotid using RANSAC and cubic splines. *Computer Methods and Programs in Biomedicine* 101(1), pp. 94-106.

Roman, K. 1980. Clinical application of an ultrasound attenuation coefficient estimation technique for liver pathology characterization. *IEEE Transactions on Biomedical Engineering* BME-27(6), pp. 312-319.

Royal College of Obstetricians and Gynaecologists. 2003. Ovarian Cysts in Postmenopausal Women, Guideline No. 34.

Rubin, J. M. 1994. Spectral doppler ultrasound. *RadioGraphics* 14(1), pp. 139-150.

Rusko, L. et al. 2009. Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images. *Medical Image Analysis* 13, pp. 871-882.

Russ, J. C. 1999. *The Image Processing Handbook*. 3rd. ed. Florida: I.E.E.E.Press.

Russell, P. ed. 1994. *Surface Epthelial-Stromal Tumors of the Ovary.* 4th. ed. New York: Springer-Verlag, pp. 705–782.

Sassone, A. et al. 1991. Transvaginal sonographic characterization of ovarian disease: Evaluation of a new scoring system to predict ovarian malignancy. *Obstetrics and Gynecology* 78(1), pp. 70-76.

Schmauder, S. et al. 2008. Cyclic changes in endometrial echotexture of cows using a computer-assisted program for the analysis of first- and second-order grey level statistics of B-Mode ultrasound images. *Animal Reproduction Science* 106(1), pp. 153-161.

Sedgewick, J. 2008. *Scientific Imaging With Photoshop*. Berkeley: New Riders, p. 301.

Sharma, M. and Singh, S. 2001. *Evaluation of texture methods for image analysis*. Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001. 18-21 Nov. 2001.

Shen, S. et al. 1997. Reproducibility of operator processing for radiation dosimetry. *Nuclear Medicine and Biology* 24(1), pp. 77-83.

Sheppard, M. A. and Shih, L. 2005. *Efficient image texture analysis and classification for prostate ultrasound diagnosis*. IEEE Computational Systems Bioinformatics Conference, 2005. 8-11 Aug. 2005.

Shi, X. et al. 2010. Detection and classification of masses in breast ultrasound images. *Digital Signal Processing* 20(3), pp. 824-836.

Shung, K. K. 2006. *Diagnostic Ultrasound : Imaging and Blood Flow Measurements*. London: Taylor & Francis, p. 215.

Simon, S. 2008. *Receiver Operating Characteristics* [Online]. Kansas City: Children's Mercy Hospitals and Clinics. Available at: http://www.childrensmercy.org/stats/ask/roc.asp [Accessed: 20 July 2010].

Singh, J. et al. 1998. Ultrasound image attributes of bovine ovarian follicles and endocrine and functional correlates. *Journal of Reproduction and Fertility* 112, pp. 19-29.

Singh, M. and Singh, S. 2002. Spatial texture analysis: a comparative study. In: *16th International Conference on Pattern Recognition, 2002.* 2002. pp. 676-679.

Sivaramakrishna, R. et al. 2002. Texture analysis of lesions in breast ultrasound images. *Computerized Medical Imaging and Graphics* 26(5), pp. 303-307.

Skates, S. J. et al. eds. 2000. *Tumor Markers in Screening for Ovarian Cancer.* New Jersey: Humana Press, p. 817.

Smutek, D. et al. 2003. Image texture analysis of sonograms in chronic inflammations of thyroid gland. *Ultrasound in Medicine & Biology* 29(11), pp. 1531-1543.

Smutek, D. et al. 2006. Automatic internal medicine diagnostics using statistical imaging methods. In: *19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006.* pp. 405-412.

Smyth, P. P. et al. 1997. Application of computer texture analysis to the Singh Index. *Br J Radiol* 70(831), pp. 242-247.

Sohaib, S. A. A. and Reznek, R. H. eds. 2002. *CT and MRI of Ovarian Cancer.* 2nd. ed. Oxford University Press, p. 427.

Spethmann, J. et al. 2010. Accuracy of computer-aided transrectal ultrasonography detection (Histoscanning) of prostate cancer in the prediction of a negative margin in radical prostatectomy patients. *European Urology Supplements* 9(2), pp. 65-66.

Srinivasan, G. N. and Shobha, G. 2008. Statistical texture analysis. In: *World Academy of Science, Engineering and Technology.* December. pp. 1264-1269.

Stoitsis, J. et al. 2006(b). A modular software system to assist interpretation of medical images: Application to vascular ultrasound images. *IEEE Transactions on Instrumentation and Measurement* 55(6), pp. 1944-1952.

Stoitsis, J. et al. 2006(a). Characterization of carotid atherosclerotic plaques using frequency-based texture analysis and bootstrap. In: *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS '06.* Aug. 30 2006-Sept. 3 2006. pp. 2392-2395.

Strub, K. L. 2005. Adult echocardiography and Doppler. *Journal of diagnostic sonography* 21, pp. 91-110.

Swinscow, T. D. V. 1997. *Statistics at Square One*. Ninth ed. BMJ Publishing Group, p. 108.

Szczypinski, P. M. et al. 2007. Mazda - a software for texture analysis. In: *International Symposium on Information Technology Convergence, 2007.* 23-24 Nov. 2007. pp. 245-249.

Szczypinski, P. M. et al. 2009. MaZda--A software package for image texture analysis. *Computer Methods and Programs in Biomedicine* 94(1), pp. 66-76.

Tan, D. S. P. et al. 2006. Mechanisms of transcoelomic metastasis in ovarian cancer. *The Lancet Oncology* 7(11), pp. 925-934.

Tesar, L. et al. 2008. Medical image analysis of 3D CT images based on extension of Haralick texture features. *Computerized Medical Imaging and Graphics* 32(6), pp. 513-520.

Testa, A. C. et al. 2005. Does quantitative analysis of three-dimensional power Doppler angiography have a role in the diagnosis of malignant pelvic solid tumors? A preliminary study. *Ultrasound in Obstetrics and Gynecology* 26(1), pp. 67-72.

Theocharakis, P. et al. 2009. Pattern recognition system for the discrimination of multiple sclerosis from cerebral microangiopathy lesions based on texture analysis of magnetic resonance images. *Magnetic Resonance Imaging* 27(3), pp. 417-422.

Thijssen, J. M. ed. 1992. *Echographic Image Processing.* Academic Press, p. 359.

Timmerman, D. 2000. Lack of standardization in gynecological ultrasonography. *Ultrasound in Obstetrics and Gynecology* 16(5), pp. 395-398.

Timmerman, D. et al. 2008. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound in Obstetrics and Gynecology* 31(6), pp. 681-690.

Timmerman, D. et al. 2000. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) group. *Ultrasound in Obstetrics and Gynecology* 16(5), pp. 500-505.

Togashi, K. 2003. Ovarian cancer: the clinical role of US, CT, and MRI. *European Radiology* 13(supplement 6), pp. L87-L104.

Tourassi, G. D. 1999. Journey toward Computer-Aided Diagnosis: Role of image texture analysis. *Radiology* 213(2), pp. 317-320.

Tsai, D.-Y. and Kojima, K. 2005. Measurements of texture features of medical images and its application to computer-aided diagnosis in cardiomyopathy. *Measurement* 37, pp. 284–292.

Tuceryan, M. and Jain, A. K. eds. 1998. *The Handbook of Pattern Recognition and Computer Vision.* 2nd. ed. World Scientific Publishing Co., pp. 207-248.

Turkington, C. and Edelson, M. 2005. *The Encyclopedia of Women's Reproductive Cancer* New York: Infobase Publishing, p. 306.

Twickler, D. M. and Moschos, E. 2010. Ultrasound and Assessment of Ovarian Cancer Risk. *Am. J. Roentgenol.* 194(2), pp. 322-329.

Valckx, F. M. J. and Thijssen, J. M. 1997. Characterization of echographic image texture by cooccurrence matrix parameters. *Ultrasound in Medicine & Biology* 23(4), pp. 559-571.

Valentin, L. 1999. High-quality gynecological ultrasound can be highly beneficial, but poor-quality gynecological ultrasound can do harm. *Ultrasound in Obstetrics and Gynecology* 13(1), pp. 1-7.

Valentin, L. et al. 2006. Ultrasound characteristics of different types of adnexal malignancies. *Gynecologic Oncology* 102, pp. 41-48.

Valentin, L. et al. 1994. Limited contribution of doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors. *Obstetrics & Gynecology* 83(3), pp. 425-433.

Van Ginneken, B. et al. 2002. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging* 21(2), pp. 139-149.

Vince, D. G. et al. 2000. Comparison of texture analysis methods for the characterization of coronary plaques in intravascular ultrasound images. *Computerized Medical Imaging and Graphics* 24, pp. 221–229.

Vine, M. F. et al. 2003. Characterization of prediagnostic symptoms among primary epithelial ovarian cancer cases and controls. *Gynecologic Oncology* 90(1), pp. 75-82.

Wang, Y. et al. 2002. Studies on tissue characterization by texture analysis with co-occurrence matrix method using ultrasonography and CT imaging. *Journal of Medical Ultrasonics* 29(4), pp. 211-223.

Weszka, J. S. et al. 1976. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics* SMC-6(4), pp. 269-285.

White, A. D. et al. 1997. Problems in the comparison of ultrasound morphology and histology of carotid artery atheromatous plaques. *Journal of Vascular Investigation* 3(4), pp. 168-176.

Whittingham, T. and Martin, K. eds. 2010. *Transducer and Beam-Forming.* 2nd. ed. Cambridge: Cambridge University Press, p. 263.

Whittingham, T. A. ed. 2000. *The acoustic output of diagnostic machines.* London: British Institute of Radiology, p. 120.

Whittingham, T. A. 2007. Medical diagnostic applications and sources. *Progress in Biophysics and Molecular Biology* 93(1-3), pp. 84-110.

Wilhjelm, J. E. et al. 1998. Quantitative analysis of ultrasound B-mode images of carotid atherosclerotic plaque: correlation with visual classification and histological examination. *IEEE Transactions on Medical Imaging* 17(6), pp. 910-922.

Wilson, W. D. et al. 2006. Sonographic Quantification of Ovarian Tumor Vascularity *J Ultrasound Med* 25, pp. 1577-1581.

Xian, G.-M. 2010. An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. *Expert Systems with Applications* 37, pp. 3737-3741.

Xiao, G. et al. 2002. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. *IEEE Transactions on Medical Imaging* 21(1), pp. 48-57.

Xu, Y. et al. 2006. Computer-aided Classification of Interstitial Lung Diseases Via MDCT: 3D Adaptive Multiple Feature Method (3D AMFM). *Academic Radiology* 13(8), pp. 969-978.

Yeh, W.-C. et al. 2003. Liver fibrosis gred classification with B-Mode ultrasound. *Ultrasound in Medicine and Biology* 29(9), pp. 1229-1235.

Zhang, J. et al. 2008. Texture analysis of multiple sclerosis: a comparative study. *Magnetic Resonance Imaging* 26(8), pp. 1160-1166.

Zhu, Y. et al. 2006. Computer technology in detection and staging of prostate carcinoma: A review. *Medical Image Analysis* (10), pp. 178 - 199.

Zimmer, Y. et al. 2003. An automatic approach for morphological analysis and malignancy evaluation of ovarian masses using B-scans. *Ultrasound in Medicine and Biology* 29(11), pp. 1561-1570.

Toshiba Xario - Specifications

Item		Specifications
Power	Line voltage	230 VAC +/-10% (for Europe) 110 to 120 VAC +/-10% (for Other 1) 220 to 240 VAC +/-10% (for Other 2)
	Line frequency	50 Hz to 60 Hz +/-1Hz
	Power consumption	Maximum 1.5 kVA
Operating environmental conditions	Ambient temperature	10^0C to 30^0C
	Relative humidity	35% to 80% (no condensation)
	Atmospheric pressure	700 hPa to 1060 hPa
Storage and transportation conditions	Ambient temperature	-10^0C to 50^0C
	Relative humidity	30% to 80% (no condensation) 50% or less if ambient temperature exceeds 40^0C
	Atmospheric pressure	700 hPa to 1060 hPa
External dimensions (not including optional units)	For system with CRT monitor	540 +/-20 mm (width) x 1455 +/-30 to 1500 +/-50 mm (height) x 814 +/-30 mm (depth)
	For system with LCD monitor	540 +/-20 mm (width) x 1360 +/-30 to 1595 +/-50 mm (height) x 814 +/-30 mm (depth)
Mass (not including optional units)	For system with CRT monitor	150 kg or less
	For system with LCD monitor	140 kg or less

-Appendix B-

Summary of the results for the influence of the ROI size on the extracted textural features.

Pixel size		GLCM	RLM	Gradient	AR	Wavelet
100 pixels	Mean	969	18	311	0.27237	3762
	Std Dev	44	1	48	0.02238	312
	CoV	4.50	6.21	15.51	8.22	8.29
400 pixels	Mean	982	65	302	0.27780	4007
	Std Dev	31	3	39	0.00687	102
	CoV	3.20	4.17	12.84	2.47	2.55
900 pixels	Mean	982	144	291	0.28012	3962
	Std Dev	17	3	23	0.00521	147
	CoV	1.72	2.31	8.04	1.86	3.71
1,600 pixels	Mean	986	250	298	0.27977	3871
	Std Dev	16	6	15	0.00494	38
	CoV	1.67	2.31	5.16	1.77	0.98
2,500 pixels	Mean	977	382	292	0.27957	3852
	Std Dev	15	9	10	0.00231	100
	CoV	1.57	2.47	3.56	0.82	2.59
3,600 pixels	Mean	971	527	285	0.28032	3978
	Std Dev	21	16	12	0.00175	57
	CoV	2.14	3.09	4.14	0.63	1.43
4,900 pixels	Mean	971	682	283	0.28144	3832
	Std Dev	24	21	9	0.00183	50
	CoV	2.43	3.04	3.15	0.65	1.31
6,400 pixels	Mean	976	842	284	0.28251	3966
	Std Dev	18	22	11	0.00151	57
	CoV	1.84	2.64	3.79	0.54	1.44
8,100 pixels	Mean	983	999	292	0.28307	3974
	Std Dev	15	29	11	0.00145	79
	CoV	1.53	2.88	3.75	0.51	1.99
10,000 pixels	Mean	971	1164	299	0.28342	3944
	Std Dev	15	30	11	0.00124	36
	CoV	1.49	2.57	3.69	0.44	0.92

-Appendix C-

Summary of the results for the influence of the ROI depth on the extracted textural features.

		Mean (SD)				
Focus (cm)	**ROI Depth (cm)**	**GLCM**	**RLM**	**Gradient**	**AR**	**Wavelet**
4	**1**	924 (34)	116 (1)	255 (35)	0.29320 (0.00411)	4,274 (264)
	2	933 (19)	116 (1)	273 (21)	0.28284 (0.00348)	4,130 (158)
	3	921 (27)	115 (1)	225 (23)	0.26702 (0.00552)	4,272 (137)
	4	933 (17)	113 (2)	246 (24)	0.26297 (0.00475)	4,297 (185)
	5	942 (32)	110 (3)	259 (50)	0.25742 (0.00810)	4,046 (162)
	6	916 (43)	103 (3)	265 (67)	0.24142 (0.00845)	4,179 (181)
6	**1**	913 (43)	116 (1)	249 (28)	0.27767 (0.00336)	4,281 (251)
	2	953 (33)	116 (1)	255 (38)	0.27598 (0.00443)	3,944 (80)
	3	939 (25)	115 (2)	248 (12)	0.26537 (0.00475)	4,058 (82)
	4	895 (34)	112 (1)	237 (24)	0.25977 (0.00495)	4,281 (158)
	5	928 (30)	109 (2)	219 (23)	0.25473 (0.00512)	4,298 (364)
	6	869 (49)	103 (8)	133 (36)	0.23919 (0.01066)	4,024 (289)
8	**1**	939 (27)	116 (2)	261 (45)	0.28134 (0.00449)	4,362 (147)
	2	928 (28)	115 (1)	252 (25)	0.28292 (0.00731)	4,228 (206)
	3	942 (21)	116 (1)	270 (25)	0.27037 (0.00532)	4,029 (112)
	4	897 (34)	112 (2)	221 (40)	0.26386 (0.00379)	4,623 (164)
	5	936 (42)	109 (1)	235 (32)	0.26032 (0.00543)	4,314 (141)
	6	953 (42)	98 (7)	195 (44)	0.25330 (0.00423)	4,182 (367)

-Appendix D-

Summary of the results for the influence of the gain setting on the extracted textural features.					
Gain (dB)	**Mean (SD)**				
	GLCM	**RLM**	**Gradient**	**AR**	**Wavelet**
80	980 (17)	429 (3)	260 (3)	0.28405 (0.00058)	3,794 (28)
85	922 (18)	441 (4)	241 (3)	0.27844 (0.00092)	3,926 (53)
90	943 (11)	446 (2)	236 (6)	0.27840 (0.00069)	3,978 (42)
95	944 (11)	453 (2)	233 (3)	0.27886 (0.00077)	3,985 (44)
100	942 (7)	462 (2)	237 (5)	0.27782 (0.00065)	4,016 (43)