THE UNIVERSITY OF BUCKINGHAM

Computer-Aided Diagnosis of Gynecological Abnormality using B-mode Ultrasound Images

By

Shan Khazendar

Department of Applied Computing
University of Buckingham
United Kingdom

A Thesis for the Degree of Doctor of Philosophy in Computer Science to the School of Science in the University of Buckingham

February, 2016

ABSTRACT

Ultrasound scan is one of the most reliable imaging for detecting/diagnosing of gynaecological abnormalities. Ultrasound imaging is widely used during pregnancy and has become central in the management of the problems of early pregnancy, particularly in miscarriage diagnosis. Also ultrasound is considered as the most important imaging modality in the evaluation of different types of ovarian tumours.

The early detection of ovarian carcinoma and miscarriage continues to be a challenging task. It mostly relies on manual examination, interpretation by gynaecologists, of the ultrasound scan images that may use morphology features extracted from the region of interest. Diagnosis depends on using certain scoring systems that have been devised over a long time. The manual diagnostic process involves multiple subjective decisions, with increased inter- and intra-observer variations which may lead to serious errors and health implications.

This thesis is devoted to developing computer-based tools that use ultrasound scan images for automatic classification of Ovarian Tumours (Benign or Malignant) and automatic detection of Miscarriage cases at early stages of pregnancy. Our intended computational tools are meant to help gynaecologists to improve accuracy of their diagnostic decisions, while serving as a tool for training radiology students/trainees on diagnosing gynaecological abnormalities. Ultimately, it is hoped that the developed techniques can be integrated into a specialised gynaecology Decision Support System.

Our approach is to deal with this problem by adopting a standard image-based pattern recognition research framework that involve the extraction of appropriate feature vector modelling of the investigated tumours, select appropriate classifiers, and test the performance of such schemes using sufficiently large and relevant datasets of ultrasound scan images. We aim to complement the automation of certain parameters that gynaecologist experts and radiologists manually determine, by image-content information attributes that may not be directly accessible without advanced image transformations. This is motivated by, and benefit from, advances in computer vision that led the emergence of a variety of image processing/analysis techniques together with recent advances in data mining and machine learning technologies.

An expert observer makes a diagnostic decision with a level of certainty, and if not entirely certain about their diagnostic decisions then often other experts' opinions are sought and may be essential for diagnosing difficult "Inconclusive cases". Here we define a quantitative

measure of confidence in decisions made by automatic diagnostic schemes, independent of accuracy of decision.

In the rest of the thesis, we report on the development of a variety of innovative diagnostic schemes and demonstrate their performances using extensive experimental work. The following is a summary of the main contributions made in this thesis.

1. Using a combination of spatial domain filters and operations as pre-processing procedures to enhance ultrasound images for both applications, namely miscarriage identification and ovarian tumour diagnosis. We show that the Non-local means filter is effective in reducing speckle noise from ultrasound images, and together with other filters we succeed in enhancing the inner border of malignant tumours and reliably segmenting the gestational sac.
2. Developing reliable automated procedures to extract several types of features to model gestational sac dimensional measurements, few of which are manually determined by radiologist and used by gynaecologists to identify miscarriage cases. We demonstrate that the corresponding automatic diagnostic schemes yield excellent accuracy when classified by the k-Nearest Neighbours.
3. Developing several local as well as global image-texture based features in the spatial as well as the frequency domains. The spatial domain features include the local versions of image histograms, first order statistical features and versions of local binary patterns. From the frequency domain, we propose a novel set of Fast Fourier Geometrical Features that encapsulates the image texture information that depends on all image pixel values. We demonstrate that each of these features define Ovarian Tumour diagnostic scheme that have relatively high power of discriminating Benign from Malignant tumours when classified by Support Vector Machine. We show that the Fast Fourier Geometrical Features are the best performing scheme achieving more than 85% accuracy.
4. Introducing a simple measure of confidence to quantify the goodness of the automatic diagnostic decision, regardless of decision accuracy, to emulate real life medical diagnostics. Experimental work in this theis demonstrate a strong link between this measure and accuracy rate, so that low level of confidence could raise an alarm.
5. Conducting sufficiently intensive investigations of fusion models of multi-feature schemes at different level. We show that feature level fusion yields degraded performance compared to all its single components, while score level fusion results in

improved results and decision level fusion of three sets of features using majority rule is slightly less successful. Using the measure of confidence is useful in resolving conflicts when two sets of features are fused at the decision level. This leads to the emergence of a *Not Sure* decision which is common in medical practice. Considering the Not Sure label is a good practice and an incentive to conduct more tests, rather than misclassification, which leads to significantly improved accuracy.

The thesis concludes with an intensive discussion on future work that would go beyond improving performance of the developed scheme to deal with the corresponding multi-class diagnostics essential for a comprehensive gynaecology Decision Support System tool as the ultimate goal.

ACKNOWLEDGMENTS

Allah the Most Gracious and Merciful

First and foremost I would like thank Almighty **GOD**, the compassionate, the almighty Merciful, who has given me the strength and the ability to complete my thesis and to make my dream come true.

This journey would not have been possible without support, guidance, and efforts of many people.

My Sponsor

I would like to thank the **Ministry of Higher Education and Scientific Research** in **Kurdistan** for offering me the Human Capacity Development Program (HCDP) scholarship to complete my PhD. Also special thanks to the University of Sulaimani for providing all supports to complete this study.

My Supervisors

I would like to express my sincere gratitude to my supervisors **Professor Sabah A. Jassim, Mr. Hongbo Du** and **Dr. Hisham Al-Assam** for their support, invaluable guidance and that I will never forget. My deepest gratitude is to **Professor Jassim** for his guidance and invaluable suggestions in computational and mathematical works, and his support from the beginning to the end of my study. I have been extremely lucky to work with him. Special thanks to **Mr. Du** for his support, guidance and advice during my study, I learned a lot from him. I'm also gratefully to **Dr. Assam** for his endless support, meaningful ideas and for his continues encouragement that has made me feel confident and overcome every difficulty I encountered, Dr. Assam was always there when I needed him.

I sincerely appreciate the support of the Head of Department, **Dr. Harin Sellahewa** and all academic staff in the Applied Computing Department /University of Buckingham.

Finally, Special thanks go to **Mrs. Sharon Salerno** Administrator of Psychology and Applied Computing Departments for her support in all respects and encouragement.

Collaborators

Many thanks also to all the collaborators involved in this work. Department of Early Pregnancy and Department of Cancer and Surgery, Imperial College, Queen Charlotte's and Chelsea Hospital in London. Department of Obstetrics and Gynaecology, University Hospital KU, Leuven, Belgium.

Firstly, many thanks for **Professor Tom Bourne,** without his help, this work would not have been successful. I am very grateful for him for offering the opportunity to work on an interesting problem, and also providing medical training courses and the datasets for this study, I will never forget his support and will always remember our first meeting. Special

thanks go to **Professor Dirk Timmerman** for providing the datasets, and for his encouragement. My deepest gratitude to **Dr. Ahmad Sayasneh** for his endless support and advice from the start to the end of this study. My special thanks to **Dr. Jessica Farren** and **Dr. Jeoren Kaijser** for their help in preparing the images and all important information related to the datasets used in this study.

My Family

Special and deepest thanks to my parents, my father **Sirwan** and to my dear mother **Maha,** I owe them everything I've achieved today. They helped me to start with this study and provided me with their constant encouragement, helpful advice, care, and affection. I would also like to thank my brother **Shad** and my sister **Shene** for their support, continuous love and good wishes whenever I needed. Special thanks to my mother-in-law, for her love and prayers. I would also like to express my gratitude to the soul of my father-in-law. Many thanks go to the rest of my family.

Last, but definitely not the least, I am greatly indebted to my beloved husband **Osman.** I find it difficult to express my appreciation enough because it is so boundless. He is my best friend and amazing husband and father. Thanks for his support and love. Without him I would never been able to complete my studies. Many thanks to my beautiful princess **Lena** and my little son **Land** for their love, patience, and understanding. They are the most precious source of my happiness.

My Friends and Colleagues

I have to thank all my friends who have supported and encouraged me during my study.

This thesis is dedicated to My Parents, Osman, Lena and Land

ABBREVIATIONS

CA-125	Cancer Antigen 125
CAD	Computer Aided Diagnosis
CRL	Crown Rump Length
DSS	Decision Support System
DSS	Decision Support System
FFGF	Fast Fourier Geometrical Features
FFT	Fast Fourier Transformation
GS	Gestational Sac
hCG	Human Chorionic Gonadotropin
k-NN	k-Nearest Neighbour
LBP	Local Binary Pattern
LR 1 & 2	Logistic Regression Models
MSD	Mean Sac Diameter
NL-Means	Non Local Means
PUV	Pregnancy Unknown Viability
ROI	Region of Interest
RMI	Risk of Malignancy Index
SVM	Support Vector Machine
US	Ultrasound
YS	Yolk Sac

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

DECLARATION

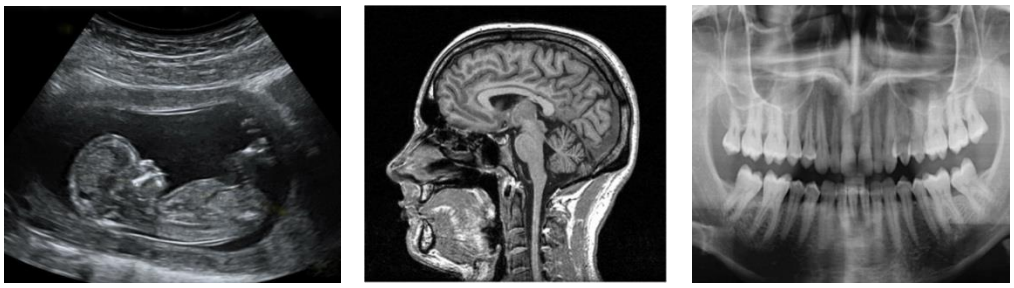
I declare that this written submission represents my ideas in my own words and where others' ideas, work or words have been included, I have adequately cited and referenced the original sources.

Shan Khazendar

CHAPTER 1

INTRODUCTION

Medical imaging refers to several technologies that are used to view body parts/organs in order to monitor, diagnose, or treat medical conditions (Bushberg and Boone 2011, WiseGeek 2013). These technologies have been increasingly deployed in the past few decades leading to significant improvement in our understanding of diseases and guiding treatments. There exist various imaging modalities such as Ultrasound (US), Magnetic Resonance (MR), X-ray, etc. They all provide a variety of images with different structures and functions of internal anatomies (Dhawan 2011, WiseGeek 2013). Figure 1.1 illustrates some examples of medical images of different modalities. However, no imaging method in existence today is capable of providing all the information that a doctor or a surgeon needs for diagnosing a condition or treating a patient (Bushberg and Boone 2011, Dhawan 2011).

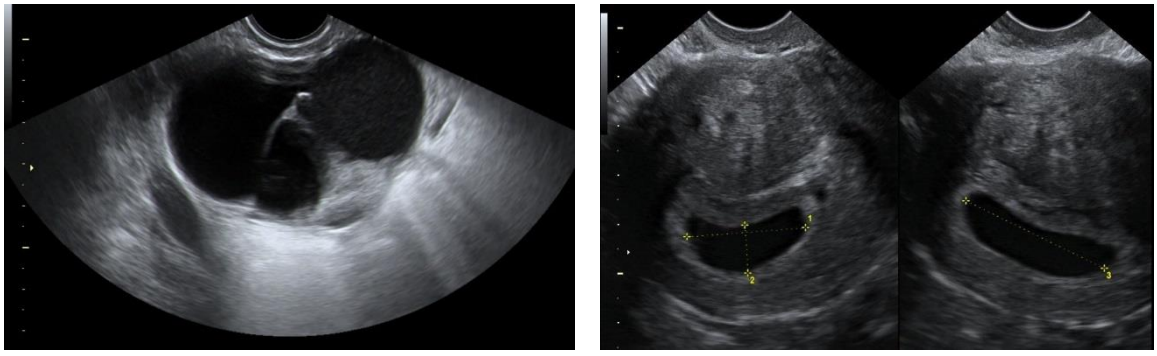


(a) (b) (c)

Figure 1. 1: Examples of different types of medical images (a) US image of fetal (b) MRI of the brain (c) X- ray of the teeth

Ultrasound imaging is probably the most commonly deployed medical image modality. It has been used for over half a century and is considered to be the most common diagnosis tool deployed in hospitals around the world, especially in detecting abnormalities in the gynaecology field (Michailovich and Tannenbaum 2006). US images are distinguished from imaging systems generated by waveforms of the electromagnetic spectrum, in that they are generated by frequency sound waves. The reflected sound wave echoes are recorded and quantised for display as a real-time image (Hangiandreou 2003). Applications of US imaging have rapidly grown from simple measurements of anatomical dimensions to detailed screening for fetal abnormalities, detection of changes in tissue texture, diagnosis of different types of tumour and detailed study of blood flow in arteries (Hoskins, Martin et al. 2010). According to (Geirsson and Busby-Earle 1991, Kinkel, Hricak et al. 2000), ultrasound is the primary imaging modality in detecting abnormalities in ovary and evaluation of the ovarian

tumours. Figure 1.2 shows a US scan image of ovarian tumour and a US scan image of the gestational sac in the early stage of pregnancy.



(a) (b)

Figure 1. 2: Examples of US scan images (a) Ultrasound image of ovarian tumour (b) Ultrasound image of Gestational sac in two different planes

Ultrasound scanning is considered as an effective imaging modality for monitoring pregnancy because of its safety without the hazard of radiation, especially in diagnosing abnormalities that lead to miscarriage at a very early stage in pregnancy (Hoskins, Martin et al. 2010). The first three months of pregnancy are the most crucial period. Monitoring pregnancy within this period enables clinics to evaluate the development, growth, and wellbeing of the fetus (Kaur and Kaur 2011). Loss of pregnancy before 24 weeks of gestation is termed as miscarriage. The estimated number of miscarriage cases in UK is 250,000 each year (England London HSCIC, 2013).

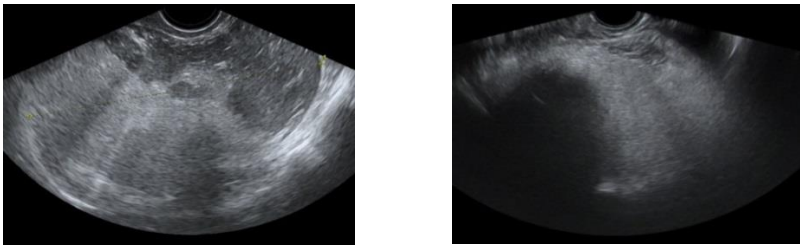
Ovarian cancer is the fifth most common cancer after breast, bowel, lung and uterus cancer and the second most common gynaecological cancer after uterus (UK 2015). Around the world, nearly 200,000 women are estimated to develop ovarian cancer every year and about 100,000 die from it. It has been reported that around 140,000 women died of ovarian cancer in the world in 2008 alone. In the United Kingdom, more than 7000 women are diagnosed with ovarian cancer each year and 4,200 deaths occurred this reported in 2014 (UK 2011). Ovarian cancer has the highest mortality rate of all gynaecologic cancers (Jeong, Outwater et al. 2000, Fishman, Cohen et al. 2005, Chan and Selman 2006, ACS 2014, UK 2015) and has been known as “the silent killer” because of its non-specific symptoms (Chan and Selman 2006). According to the first prospective study in (Braem, Onland-Moret et al. 2012), multiple miscarriages are also associated with an increased risk of ovarian cancer.

This thesis is concerned with the analysis of ultrasound images for the detection and classification of gynaecological abnormalities. In particular, the focus is on analysing US

scan images of ovarian tumour for signs of malignancy as well as the analysis of US scan images of gestational sac for diagnosing miscarriages. Our investigations will follow a pattern recognition approach and develop digital image processing and analysis techniques both in the spatial domain and frequency domain. We also test the performance of our developed schemes.

Automatic medical image analysis and their use in detecting/classifying abnormalities is shrouded with tough challenges for a variety of reasons such as inadequate image quality as well as shortcomings of the existing computational models. For example in medical diagnosis, distinguishing between positive and negative cases is a crucial task, which gives rise to two types of errors: false negative and false positive. A false-positive error occurs when test result indicates that a person has a specific disease or condition when the person actually does not. A false negative error occurs when test results indicate a person does not have a disease or condition when the person actually has it. In both situations there could be serious consequences and inconveniences. For instance, a patient that has been diagnosed falsely as having malignant tumour (false-positive) may undergo unnecessary procedures including surgery and/or biopsies (Myers, Bastian et al. 2006). On the other hand, if a patient has been falsely diagnosed with benign tumour, vital timely treatments of malignancy may be missed with the severe consequence on fatality and survival rates.

In general biomedical image analysis, a specific region in the image is relevant to the purpose of the medical examination, which is referred to a region of interest (ROI). For example in US images of ovarian tumour, the ROI is the tumour lump while the ROI of US scan for pregnancy tests is the gestational sac. Detecting the ROI automatically with reasonable accuracy is a serious challenge for a variety of reasons including inadequate image quality as well as the absence of a crisp universally agreed model of these regions. The following US images illustrate such difficulties.



(a) (b)

Figure 1. 3: Examples of challenging cases for automatic segmentation of ovarian tumour (a) Tumour and its background have a similar texture (b) A tumour with unclear border

Currently, quantifying the region of interest and discriminating between positive and negative cases in the areas of miscarriage diagnosis and ovarian tumour are done manually from real-time scanned ultrasound images. The diagnostic procedure entirely depends on the operator experience and involves multiple subjective decisions. Subjective decision-making can result in inter- and intra-observer variations. Inter-observer variation is the amount of difference between the results obtained by two or more observers when examining the same material. Intra-observer variation is the amount of difference one observer experiences when observing the same material more than once. Such differences may increase difficulties and even lead to errors in the diagnosis stage, which may have undesired consequences on patient health and increased patient anxiety (Pexsters, Daemen et al. 2010). Moreover, according to (Wang, Itoh et al. 2002, Huang, Chen et al. 2008, Rocha, Campilho et al. 2011), interpretation of an ultrasound image is highly dependent on the ability and experience of the medical observer as well as the reliability of the data preparation.

In order to avoid human errors in both the quantification stage and diagnosis stage, as well as reduce the false negative and false positive rates, computer-based image processing and analysis tools need to be developed, tested and integrated into the clinical diagnosis process. Addressing such a challenging task in the analysis of ultrasound images for the detection and classification of gynaecological abnormalities, or for that matter in any area of biomedical image analysis, rely on serious research investigations and development of interdisciplinary technology that combines image processing techniques, machine learning and pattern recognition solutions in close consultation with the domain experts. We should recognise that computer-based tools can never compensate for the wealth of knowledge and expertise accumulated by medical experts and specialists through long life training. The hybrid approach that combines computing tools and human experts is the reliable strategy for achieving improved accuracy of abnormality detection, patient care and patient management. In addition, such a technology can provide tools that will serve the purpose of training radiology students and personnel in medical schools and hospitals.

In general, when designing computer-based tools for medical image analysis, one needs to be clear about the purpose that the system is serving. If the tool is not to be endowed with decision-making capabilities, then the required tool could be designed to work as a traditional software tool to just automate the extraction of the features/parameters that the domain expert requires. On the other hand, another type of computer-based tools is needed when decision capabilities are expected. Such tools must go beyond traditional expert systems and deploy

some machine learning techniques that may need to extract features not normally used by the experts. Designing computerised tools with decision capabilities is desirable but much more challenging and success depends on selecting the appropriate data representation and data analysis models. This thesis is concerned with the development of tools that have characteristics of the second type of computerised tools while the role of domain experts continue in the evolution of the tool in the future, i.e. the tools will continue to be guided by the expert knowledge and needs.

1.1 Thesis Aim, Objectives and Contributions

1.1.1 Thesis Aim and Objectives

This thesis is devoted to the analysis of ultrasound images, obtained for gynaecological examinations, either during routine pregnancy scans or for diagnostic purposes in relation to ovarian abnormalities. The overall aim of this thesis is to develop, and test the performance of novel automated and computerised solutions that analyse gynaecological ultrasound images to detect and classify abnormal objects or cases that have health implications for women. Advances in imaging technology, image processing/analysis theory and the emergence of sophisticated data mining and machine learning models provide the appropriate incentive for our computational investigations.

In particular, the main objectives of this thesis include investigating automatic methods for discriminating between different types of ovarian tumours (Benign or Malignant), developing automatic methods to analyse ultrasound images of the gestational sac GS, in order to identify miscarriage from Pregnancy Unknown Viability (PUV) cases. A suite of the intended computerised methods can form the core of Computer Aided Diagnosis (CAD) tools that could provide support for biomedical scientists to achieve efficiency as well as reducing false negative and false positive cases.

Reliable CAD tools for the main applications, under investigation in this thesis, are expected to help in: (1) diagnosis of miscarriage in the very early stages of pregnancy, (i.e. discriminating between Miscarriage and normal/PUV cases); and (2) distinguishing between different types of ovarian tumours (Benign/Malignant).

Currently, the first task is dealt with by manually measuring the size of the Gestational Sac GS from live ultrasound scan images. Computerising this task is very much dependent on accurate detection of the GS in the scanned US images, and knowledge of known facts about

the geometry of the GS to determine its dimensions. In theory, the developed tool can work as *Expert-Guided CAD (EG-CAD) tool* and diagnosing miscarriage should be based on some thresholds specified by the domain expert community. However, in the course of this work and the various discussions with domain experts we come to realise that different medical communities apply different criteria for declaring miscarriage (see Chapter 2 for more details). This an incentive to investigate some machine learning training techniques that can learn from the variety of datasets of pregnancy US scan images. The results of experiments conducted to test the performance of the developed tool and procedures, that include subjective tests, demonstrate the added value of going beyond the development of EG-CAD tool and confirm the benefits of using data mining techniques for intelligent CAD tools.

Designing CAD tools to computerise ovarian tumour diagnosis is very much more challenging than dealing with miscarriage detection. This due to many factors; the most important factor is the difficulty of encapsulating domain-experts knowledge using simple rules/heuristics. Although tumour diagnosis/classification by medical experts rely on examining the ovarian US image and take into account measurements/features visible and recognisable by them from the tumour images, such as texture and size of lesion, they conduct a variety of medical tests and their final decisions are made by a team of experts who rely on technical knowledge acquired through intensive life-long training. Therefore, our approach to deal with this challenge by adopting machine learning strategy at the outset whereby a variety of ovarian US image dataset will be investigated, analysed and computerised tools will be developed through training based on a hybrid set of image features that combines texture features in the spatial and the frequency domains. These texture features used in the developed tools from a mathematical model of the image texture information that are *estimated* and *categorised* by medical experts using their knowledge gained over time. The most important adopted features are either difficult to be determined by human visual examination or even not accessible directly by medics e.g. frequency domain image coefficients. Our investigations that result in the development of *Machine Learning* based CAD tools will incorporate different classifiers; test their performances and that of the fusion at different levels.

The first step in the process of developing automatic analysis methods and tools for the intended application is the detection of the region(s) of interest (ROI) and removal of the background. In this research project, the experimental datasets of US images were acquired from different machine settings and sources. Moreover, the operators of the equipment may

have a different level of experience. Furthermore, achieving reliable segmentation of ROI depends on the removal of noise and image artefacts that may affect US images. Accordingly, automatic analysis methods start by pre-processing the US scan images for enhanced image quality and reduced noise level in order to make objects of interest more identifiable and well prepared for the ROI segmentation. Beside the sought after automatic analysis, the enhanced images output from such pre-processing procedures can also be used directly in the manual process of diagnosis by experts.

This thesis also aims to perform quantification and data analysis following feature extraction using computational techniques to detect interesting textural and anatomical changes inside the tumours or gestational sacs. These extracted features then can be used as a key to distinguish between different classes in each of the two main applications under consideration in this thesis. Furthermore, to develop automatic methods that can be closer to the reality, we introduced the idea of supporting the classification with a level of confidence. This automatic method provides the experts with a decision of a certain degree of strength in classifying a particular case. These following points summarised the main objectives of our work:

- De-noising and enhancing the ultrasound scan images used in the intended applications.
- Automatic segmentation of the region of interest ROI in US images.
- Defining and identifying features from greyscale (B-mode) ultrasound image that is deemed to have an impact on the diagnostic processes. These features extend the list of the features that are currently extracted manually by experts, to include new image features that relate to different mathematical models of images.
- Developing automated procedures to extract and quantify features from the extracted ROI, and test their reliability.
- Implement appropriate classification tools with an acceptable level of accuracy.
- Develop CAD tools that incorporate the above objectives and are endowed with decisions recommendations. These tools should be designed in a manner that supports the clinician through the decision- making process in the diagnosis of ovarian tumours and miscarriage diagnosis.
- Develop a framework for evaluating the achieved accuracy of diagnostic decision output by the above tools. Decisions are to be recommended with a level of confidence in the decision. Such a confidence based prediction outcomes are to be

more meaningful and useful for reducing false positives and false negatives in the diagnosis of the ovarian tumour as well as in the area of miscarriage cases.

1.1.2 Thesis Contributions

The research carried out in this thesis has made a number of contributions:

- Developing an automatic method for enhancing ultrasound images of ovarian tumours based on Non-local mean filter followed by a negative transformation and the absolute difference function. This enhancement method highlights the anatomical information, especially visualising the inner border inside the malignant tumours, this tool can be easily used in the clinic.
- Developing an automatic enhancing method of the ultrasound image of the gestational sac based on a mean related operation.
- Automatic segmentation of ultrasound image of the gestational sac based on Otsu thresholding followed by a set of image processing procedures to remove all non-sac binary objects.
- Investigating the effectiveness of local features in comparison to global features for gynaecological US images.
- Proposing a set of novel frequency domain features to be extracted from binarized FFT spectrum.
- Automating the quantification of known features of US images of the GS to identify miscarriage cases at early stages of pregnancy. We also investigated and evaluated a range of other geometric features of the GS for miscarriage detection.
- Investigating the contribution of image texture features in spatial/frequency domains of US images to discriminate different classes of abnormalities for both applications.
- Automatic classification of ovarian tumour. Furthermore, supporting the classification accuracy with the level of confidence, and introducing the idea of "Inconclusive case" in classification task for those cases that the machine is not sure about the class label, which is based on levels of confidence in a feature-based decision level fusion scheme.
- Automatic classification of the gestational sac. Further, supporting the classification accuracy with the level of confidence.

1.2 Research Collaborators

The datasets of this study were provided by our collaborators with all required information. The ethical approval in using the data was granted by the School of Science Ethics Committee. We are grateful for our collaborators for offering the advanced training course and providing useful information regarding the medical side of the images. The collaborators are acknowledged as follows:

- Department of Early Pregnancy, Imperial College, Queen Charlotte's and Chelsea Hospital, London, UK.
- Department of Cancer and Surgery, Queen Charlotte's and Chelsea Hospital, Imperial College, London, UK
- Department of Obstetrics and Gynaecology, University Hospital KU Leuven, Leuven, Belgium.

1.3 Training Courses and Invited Talk

1.3.1 Training Courses

- International Ovarian Tumour Analysis (IOTA) group, International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) educational course, **Advanced Course in Gynaecological Ultrasound: Imaging in Oncology**, on 17-18 January 2014, London, UK.
- International Ovarian Tumour Analysis (IOTA) group, International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) educational course, **Advanced Course in Gynaecological Ultrasound: Using Ultrasound for the Diagnosis and Management of Gynaecological Malignancy**, on 30 November and 1 December 2012, London, UK.

1.3.2 Invited Talk

- International Ovarian Tumour Analysis (IOTA) group, International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) educational course, **Advanced Course in Gynaecological Ultrasound: Imaging in Oncology**, on 17-18 January 2014, London, UK.
- First International Ovarian Tumour Analysis (IOTA) Congress, Department of Obstetrics and Gynaecology, University hospital KU Leuven, on 26-27 April 2013, Leuven, Belgium.

1.4 Publications and Presentations

1.4.1 Publications

i. **Khazendar S.**, Sayasneh A., Al-Assam H., Du H., Kaijser J., Ferrara L., Timmerman D., Jassim S., Bourne T., **Automated Characterisation of Ultrasound Images of Ovarian Tumours: The Diagnostic Accuracy of a Support Vector Machine and Image Processing with a Local Binary Pattern Operator**, Facts, views & vision in ObGyn, 2015, Vol. 7, No.1, pp. 7-15.

ii. **Khazendar S.**,Farren J., Al-Assam H., Sayasneh A., Du H., Jassim S., Bourne T., **Automatic Identification of Miscarriage Cases Supported by Decision Strength using Ultrasound Images of the Gestational Sac**, Annals of the BMVA, UK, 2015, No. 5, (pp. 1−16).

iii. **Khazendar S.**, Sayasneh A., Al-Assam H., Du H., Kaijser J., Ferrara L., Timmerman D., Jassim S., Bourne T., **Automated Classification of Static Ultrasound Images of Ovarian Tumours based on Decision Level Fusion**, 6th Computer Science and Electronic Engineering Conference (CEEC), University of Essex, UK, 2014, (pp. 148-153), IEEE.

iv. **Khazendar S.**,Farren J., Al-Assam H., Sayasneh A., Du H., Jassim S., Bourne T., **Automatic Identification of Early Miscarriage Based on Multiple Features Extracted From Ultrasound Images**, Medical Image Understanding and Analysis Conference MIUA, City University London, UK, 2014, (pp.131-136).

v. **Khazendar S.**,Farren J., Al-Assam H., Sayasneh A., Du H., Jassim S., Bourne T., **Automatic Segmentation and Classification of Gestational Sac Based on Mean Sac Diameter using Medical Ultrasound Image**, SPIE Sensing Technology+ Applications. International Society for Optics and Photonics, Baltimore, Maryland, USA, 2014, vol. 9120, (pp. 91200A-91200A).

1.4.2 Posters

i. **Khazendar S.**, Sayasneh A., Du H., Timmerman D., Bourne T., Preisler J., Guha S., Kaijser J., Jassim S., **Enhancing Static Ultrasound Images for Classification of Types of Human Ovarian Tumour**, Biotrinity Meeting, , Newbury, Berkshire, UK, 2013.

ii. **Khazendar S.**, Sayasneh A., Du H., Timmerman D., Bourne T., Preisler J., Kaijser J., Jassim S., **A Preliminary Study of Automatic Classification of Ultrasound Images of Ovarian Tumours using Support Vector Machine classifier**, Human Capacity

Development Program HCDP, Kurdistan students conference, University of Nottingham, Nottingham, UK, 2013.

1.5 Dissertation Layout

The dissertation layout will be as follows:

Chapter 2: Presents background information about medical images, ultrasound scan images, and female reproductive system with detailed information about ovarian tumours and miscarriage cases.

Chapter 3: This chapter briefly describes our research methodology and highlights the research process. The chapter also explains the experimental protocol and details about datasets that are used in this study.

Chapter 4: Reviews the literature on de-noising, enhancement and segmentation of ROI in ultrasound images. The knowledge acquired will be exploited to initiate our first research component of this thesis, by developing and proposing methods for enhancing ultrasound images of ovarian tumours as well as our method for automatic segmenting of the gestational sac.

Chapter 5: Presents a literature review of features extraction of ultrasound images as a prelude to present our proposed novel methods for feature extraction. We shall also present results of experiments to test the effectiveness of those extracted features in diagnosing the ovarian tumour status and miscarriage cases.

Chapter 6: This chapter is divided into three parts. The first part starts with the literature review on classification models. We conduct and discuss the results of experiments conducted on the appropriate US datasets to test the performance of using another different classifier that different from classifier method that used in Chapter 5. In the second part, we introduce our novel idea for adding the level of the confidence to the accuracy of classification. The concept of fusion in different levels is introduced in the third part. Moreover, all experiments, results and discussion are presented in this chapter.

Chapter 7: Concludes the thesis with a summary of the major findings from this research and highlights potential future research directions.

CHAPTER 2

BACKGROUNDS

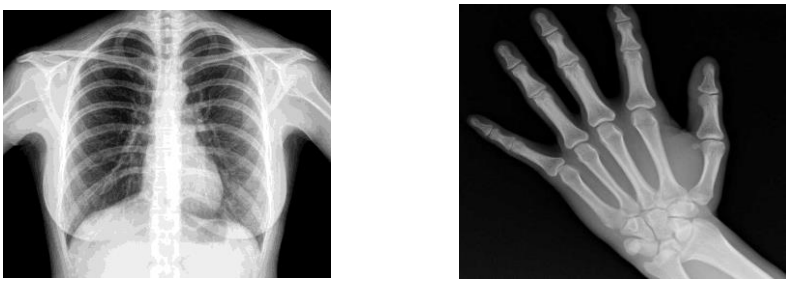
The field of medical imaging and their use as an aid in diagnosis and treatment has evolved and expanded rapidly in the early decades of last century. This chapter is devoted to describing the basic background of medical imaging and the essential medical concept of miscarriage and ovarian cancer. This chapter highlights the current clinical techniques used to characterise ovarian tumours and detect miscarriage cases. We briefly introduce various medical imaging systems and in particular, describe the different types of ultrasound imaging systems and their importance in detecting and diagnosing gynaecological abnormalities and diseases.

This chapter is divided into three main sections. Section 2.1 describes the aims of different medical imaging systems in abnormality detections. In particular, we introduce in detail ultrasound imaging systems as the ideal imaging system for gynaecological disease diagnoses. We also outline the process of ultrasound image acquisition. Section 2.2 describes the female reproductive system, highlighting key stages of pregnancy development. In section 2.3, we provide an overview of different types of ovarian tumours. In particular, we review all existing rules and models for ovarian cancer diagnoses that are currently in use. Section 2.4 gives a background of the normal pregnancy and miscarriage, and introduces the current rules in the diagnosis of abnormality in early stages of pregnancy based on the ultrasound image variables. Section 2.5 will be the brief summary of this chapter.

2.1 Medical Imaging Systems

Medical imaging refers to a number of different technologies that are used to view the human body in order to monitor, diagnose, or treat medical conditions (Bushberg and Boone 2011, Dhawan 2011) The main goal is to provide a picture of internal body structures in a way which is as non-invasive as possible (WiseGeek 2013). For example, the medical imaging system can help in determining bone fractures, detect and classify tumours and monitor pregnancy stages, etc. There are a variety of methods to obtain the images, depending on the technology being used and the area of the body which is being imaged (WiseGeek 2013). Common types of medical imaging modalities include:

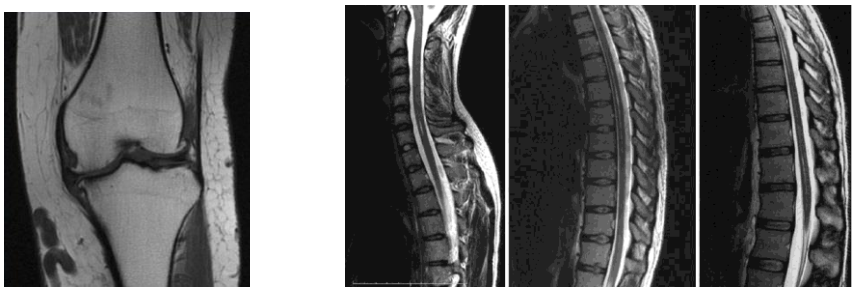
- X-ray. This imaging modality is a type of radiation; it has a higher frequency and can pass through the human body. The parts of the body that are made up of dense material, such as bones, appear as clear white areas on an X-ray image, The parts of the body that are made of softer tissues, such as heart and lungs, show up as darker areas. This modality is normally used to create images of bone fractures, tooth problems, abnormal curvature of the spine, lung problems, heart problems, and breast cancer. Because of the high levels of radiation, there is slightly increased the risk of cancer (Suetens 2009). Figure 2.1 shows two X-ray images.



(a) (b)

Figure 2. 1: Example of X-ray images (a) Chest (b) Hand

- Magnetic Resonance imaging (MRI). This modality uses magnetic fields and radio waves to show detailed images of organs, soft tissues, bones, ligaments and cartilage (Suetens 2009). During the scan, the patient lies inside a tube. MRI can be used for brain and spinal cord, bones and joints, breasts, heart, blood vessels and internal organs, such as the lungs and liver. The disadvantage of MRI is that some patients may find it claustrophobic. Also, MRI cannot be undertaken in some situations (e.g. when there is a pacemaker). Figure 2.2 shows examples of MRI images.

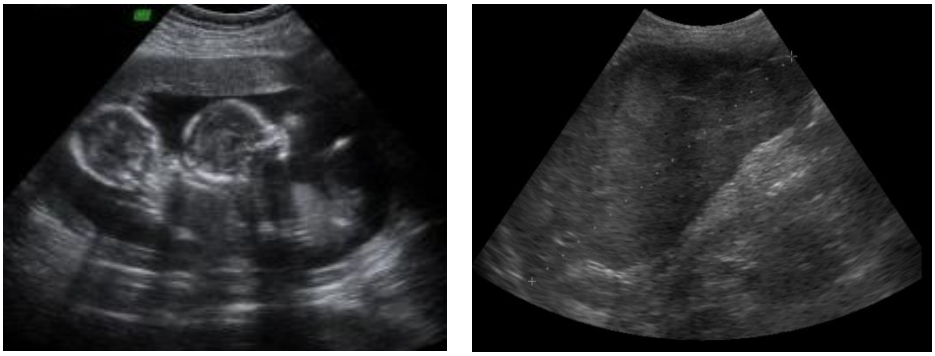


(a) (b)

Figure 2. 2: Examples MRI images (a) Knee (b) Right to Left: cervical, thoracic and lumbar spines

- Ultrasound Scan (US). An ultrasound is a “cyclic sound pressure with a frequency greater than the upper limit of human hearing (20 kHz)”. Ultrasound scan is used to create images of soft tissue structures, such as the gallbladder, liver, kidneys, pancreas, bladder, and other organs and parts of the body. Ultrasound can also

measure the blood flow inside vessels, monitoring fetal growth and detecting tumours (Suetens 2009). Figure 2.3 shows examples of US images. The main disadvantages of US it is highly operator dependent in (1) obtaining high-quality images and (2) interpreting these images. The existence of air in the body is also one of the factors which can affect image quality and the body mass index BMI (the person body size). A special preparation may be required before a procedure (e.g. full or empty bladder). Figure 2.3 shows examples of US images.



(a) (b)

Figure 2. 3: Examples of US images (a) 20 week twins (b) Normal liver

While each type of imaging modality has its different use, it is not always the case that one kind of imaging is better than another. Which type of the medical imaging system to use depends on the medical conditions and diagnosis of the diseases. Among medical imaging modalities, Ultrasound image is currently one of the most important, widely used, and multipurpose imaging modalities in medicine. The most common use of ultrasound imaging system is in detecting gynaecological abnormalities (Geirsson and Busby-Earle 1991).

2.1.1 Medical Ultrasound Image

An Ultrasound Scan (US) is sometimes called a Sonogram. It has been in use since the mid-twentieth century. Karl Theo Dussik , an Austrian neurologist, was the first to use ultrasound as a medical diagnostic tool to image the brain (Edler and Lindström 2004). Today, US is one of the most commonly used imaging technologies in medicine, especially in gynaecological disease detection because ultrasound imaging is considered to be safe without ionising radiation, non-invasive, portable, and relatively inexpensive in cost when compared with other imaging modalities, such as MRI and CT (Geirsson and Busby-Earle 1991).

Furthermore, the ultrasound scan is a real-time imaging system. US images are also tomographic, i.e. offering a “cross-sectional” view of anatomical structures. An ultrasound

scan can be used in several different ways, such as monitoring an unborn baby, diagnosing a condition such as an ovarian tumour or guiding a surgeon during certain procedures (Chan and Perlas 2011).

2.1.1.1 Ultrasound Equipment and the Process of Scanning

An ultrasound scanner consists of a number of components including a transducer (probe), a central processing unit (CPU), display screen, keyboard/cursor, disc storage devices and a printer. The transducer is a small hand-held device, and come in different shapes and sizes for using in different scanning purpose (Hoskins, Martin et al. 2010). The probe sends out frequency sound waves into the body and then listens for the returning echoes from the tissues in the body. The ultrasound image is directly visible on a video display screen (monitor) (Hoskins, Martin et al. 2010). "The image is created based on the amplitude (strength), frequency and time it takes for the sound signal to return from the area of the patient being examined to the transducer and the type of body structure the sound travels through". There are different kinds of ultrasound machine ranging from very large machines which are fixed in special clinical rooms, to small portable lightweight machines that are mobile and can be carried by a sonographer. Figure 2.4 shows three different types of the ultrasound machine.

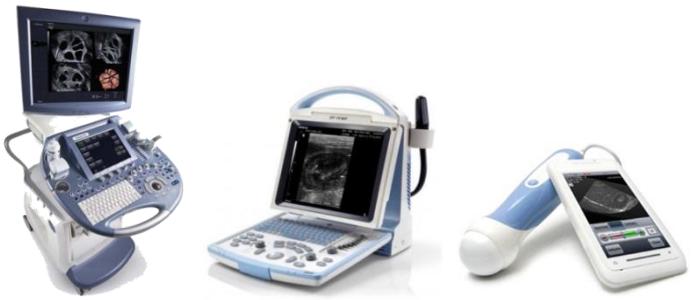


Figure 2. 4: Examples of three different sizes of ultrasound machines [12]

The scanning process starts by applying a thin layer of jelly that is placed between the probe and the skin to make sure all the sound enters the body. The probe contains a transmitter and a receiver. A pulse of ultrasound is sent out by the transmitter. The pulse is reflected from a surface and returns to the receiver. The ultrasound machine measures how long it takes for the pulse to return (Hoskins, Martin et al. 2010, NHS 2015). Figure 2.5 shows the functional components of a typical ultrasound machine and illustrates how the process of scanning works.

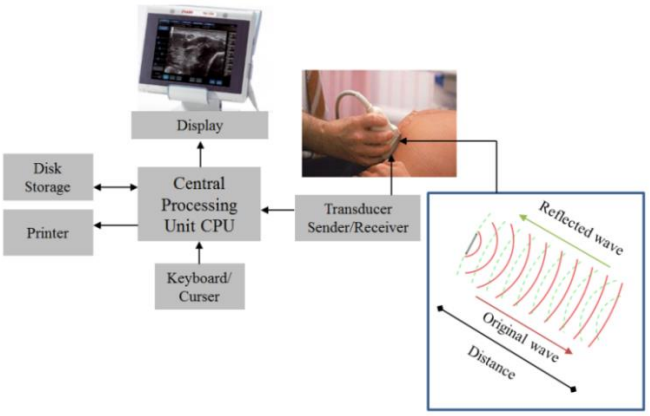
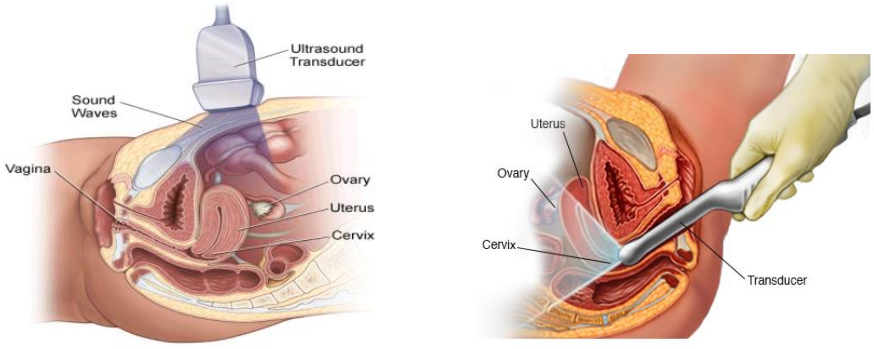


Figure 2. 5: Overview of scanning process

In gynaecology, the most common two methods of ultrasound scan are the transabdominal and transvaginal scans which differ in shapes of the probe due to its application to the different parts of the body (NHS 2015):

1. Transabdominal ultrasound (external ultrasound). A transducer probe is located onto the patient skin surface and moved over the part of the body being scanned. See Figure 2.6(a).

2. Transvaginal ultrasound (internal ultrasound). A specially shaped transducer is placed inside a woman's vagina. A transvaginal ultrasound is ideal for diagnosis of female problems such as fertility, early pregnancy scanning, or closely search for a certain condition such as an ovarian tumour or uterus. See Figure 2.6 (b).



(a) (b)

Figure 2. 6: Different types of ultrasound scan (a) External transabdominal scanning (b) Transvaginal scanning

2.1.1.2 Types of Ultrasound Images

In general, there are four types of ultrasound images that may be generated by the different equipment. Using which type depends on the condition to be diagnosed and the purpose of scanning (Palmer 1995).

1. B-mode (Brightness-mode) Ultrasound Imaging

The B-mode is a most common type of ultrasound image. It is a sequence of two-dimensional cross section images of the scanned part. The real-time scanning it can be taken in different planes as explained in below and illustrated in Figure 2.7

- The sagittal plane. It divides the body into left and right.
- The coronal plane or frontal plane. It divides the body into back and front portions.
- The transverse plane. It divides the body into cranial and caudal (head and tail) portions.

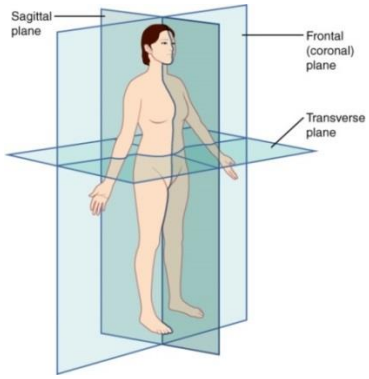
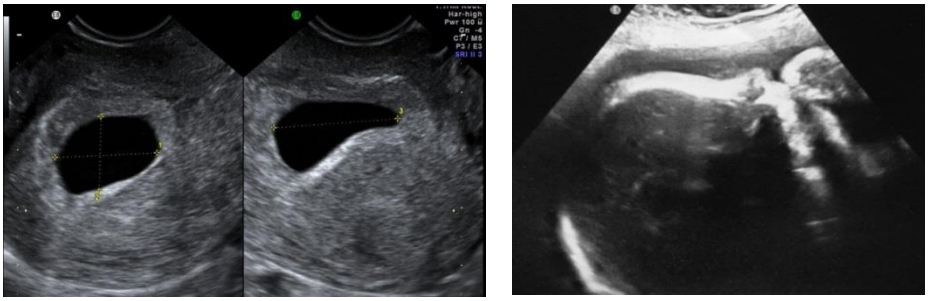


Figure 2. 7: Illustrates the transverse, sagittal, and coronal planes of the body

Gynaecologists usually examine scan image in the sagittal plane, i.e. the vertical longitudinal plane to locate the ROI, where two points on the boundaries of the ROI are selected using calibre marks to measure its diameters (major and minor). Afterwards, they change the probe angle by 90 degrees to capture the image in the transverse plane, i.e. the horizontal plane that is perpendicular to the coronal and sagittal planes, and then takes another diameter measurement; these measures depend on the purpose of diagnosis. Figure 2.8 shows examples of 2D B-mode ultrasound images.



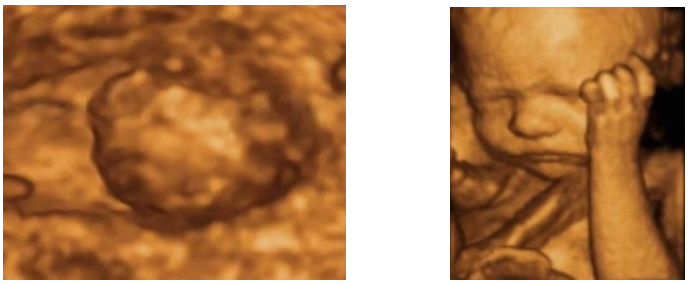
(a) (b)

Figure 2. 8: Examples of 2D B-mode ultrasound images (a) US in Sagittal and Transvers plane for GS (b) US image in Sagittal plane of 37 weeks of fetal

2. 3D Ultrasound Imaging

This is performed by scanning tissue cross sections at several different angles and rebuilding the data received into a three-dimensional image (Fenster and Downey 1996). Figure 2.9 illustrates two 3D ultrasound images. The 3D images allow observers to get a better appearance of the:

- Early detection of cancerous and differentiation between benign tumours.
- Visualising a fetus development, especially for detecting the abnormal development of the face and head.
- Visualizing blood flows in various organs or a fetus.



(a) (b)

Figure 2. 9: Examples of 3D ultrasound images (a) 3D US images of ovarian tumour (b) 3D US images of fetal

3. 4D Ultrasound Imaging

In the 4D ultrasound, the fourth time dimension allows a 3-dimensional picture in real-time and creates the most accurate representation. In some cases, 3D and 4D ultrasound image can detect abnormalities not seen in 2D ultrasound (Smythe 2004).

4. Doppler Ultrasound Imaging

Techniques of Doppler ultrasonography have been available to clinicians for nearly 40 years (Boote 2003). It helps doctors evaluating blood flow through major arteries and veins. Three types of Doppler ultrasound are currently also used in addition to 2D grey scale ultrasound imaging: Colour Doppler, Power Doppler, and Spectral Doppler (Palmer 1995). Figure 2.10 presents some examples of 2D images ultrasound image with and without colour Doppler. In Figure 2.10 (d), the image shows a very high vascularization for a malignant ovarian tumour.

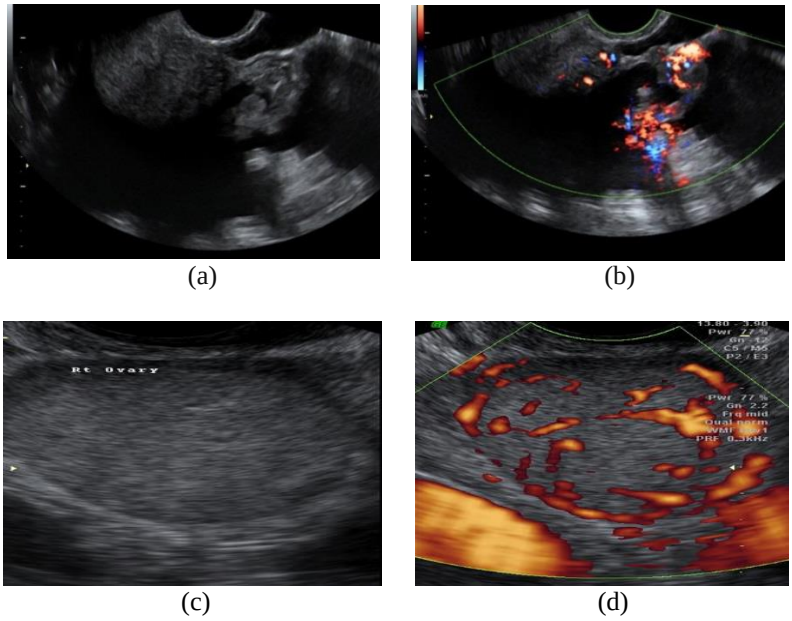


Figure 2. 10: Examples of 2D US images of colour Doppler (a) US image of ovarian tumours (b) Colour Doppler US image of ovarian tumour (c) US image of ovarian tumours (d) Colour Doppler US image of ovarian tumour

2.1.1.3 Common indications of ultrasound scan in Obstetrics and Gynaecology

Here is a short list of the most indications of ultrasound scanning Obstetrics and Gynaecology (Mason, Broaddus et al. 2010):

- To diagnose early miscarriage by measuring gestational sac in the very early stage of pregnancy.
- To estimate the expected date of delivery by measuring the fetus size.
- To determine the position of the fetus (cephalic, head down and breech).
- To check the position of the placenta, to avoid bleeding during birth.
- To determine the number of fetuses in the uterus.
- To identify the fetus gender.
- To assess the fetus growth.
- To detect ectopic pregnancy (outside the uterus pregnancy).
- To evaluate the amniotic fluid.
- To diagnose gynaecological tumours
- To characterise ovarian tumours.

Moreover, US are also used in Cardiology, Urology and ultrasound scans guide a surgeon performing some types of biopsy.

Like other image modalities, Ultrasound scan images have weakness/limitations during the use (Mason, Broaddus et al. 2010):

- Ultrasound waves cannot penetrate through bones, and are affected by environmental factors e.g. air or gas, i.e. US images may not be able to produce clear and detailed images of some parts of the body. For example, US modality is not used to scan the brain, because it is surrounded by bone.
- Large patients are harder to image by ultrasound because the greater amount of tissue decreases the sound ability to penetrate into the body.
- Ultrasound images contain noises such as speckle noise which result in reduced contrast in the images; this may affect the diagnosis accuracy.

2.1.2 Other Imaging Modalities used in Gynaecology

Computed tomography (CT) has been used mostly on patients with gynaecological malignancies to assess the extension and separate of the disease. However, ultrasound is at least as good as in characterising ovarian tumours, comparing with CT. Besides being more expensive, another major disadvantage of CT imaging is that it contains ionising radiation. According to (Chen, Ruiz et al. 2003, Jokubkiene, Sladkevicius et al. 2007, Alcázar and Jurado 2011), a combination of ultrasound morphology and Doppler blood flow is comparable to MRI. However, the major limitations of MRI are: (a) not available in every clinical centre, (b) more expensive comparing with ultrasound image and (c) more time-consuming compared with the US. MRI is recommended when US findings is not diagnostic (Sayasneh, Wynants et al. 2013). According to the Royal College of Obstetricians and Gynaecologists RCOG guideline (RCOG 2003) there is no routine role yet for Doppler, MRI or CT in ovarian tumours characterization due to the relative expense and limited availability of these modalities, which can cause a delay in the clinical management.

2.2 Background on Female Reproductive System

The female reproductive system contains two main parts: uterus and ovaries. The female reproductive organs consist anatomically of five parts (Hamlett and Koob 1999). Figure 2.11 illustrates each of the five parts and how they are connected.

1. Uterus. It is a pear shape and it normally hosts the fetus. The uterus size changes based on age and pregnancy status.
2. Cervix. It is the lowest narrow portion of the uterus where it joins with the top end of the vagina.

3. Vagina. It is a flexible muscular tube that connects the cervix of the uterus to the external of the body.
4. Fallopian tubes. Are a couple of muscular tubes that extend from the left and right corners of the uterus to the edge of the ovaries
5. Ovaries. are a pair of oval-shaped structuers, measuring 2-4 cm in diameter and 1 cm thick. They are located within the pelvis, one on each side of the uterus. Two different functions of the ovaries are: (i) to produce the eggs, and (ii) to produce the female hormones. Figure 2.12 represents an ultrasound image of a normal ovary with its corresponding laparoscopic view.

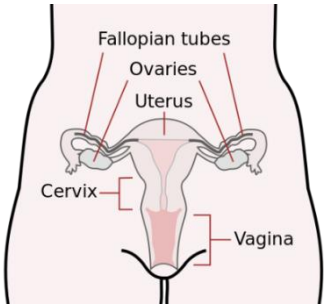
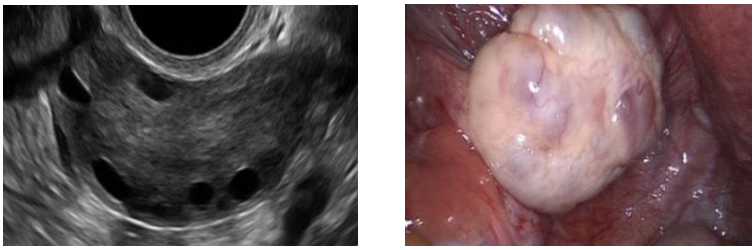


Figure 2. 11: Schematic drawing of female reproductive organs, frontal view



(a) (b)

Figure 2. 12: Represents ultrasound image of a normal ovary with its corresponding laparoscopic view (a) Ultrasound image of the normal left ovary (b) Laparoscopic view of the normal left ovary

The ovarian follicle is a cellular structure created in the ovary where immature eggs mature, leading to ovulation and either pregnancy or menstruation (Krivanek and Sonka 1998, Hamlett and Koob 1999). Women are born with nearly a million ovarian follicles, each with the possibility to grow into an egg with the chance for fertilisation. Figure 2.13 shows a normal ovary and the stages of ovulation.

Furthermore, ovaries produce the female sex hormones, oestrogen and progesterone, the hormones control the menstrual cycle. As women get older (postmenopausal), these hormones and periods eventually stop altogether.

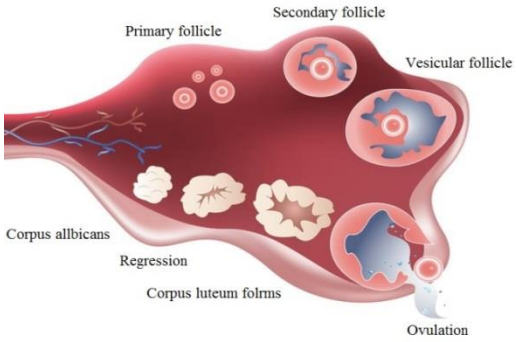


Figure 2. 13: Semi diagrammatic representation of a normal ovary and the stages of ovulation (Tatjana-Mihaela 2002)

In this thesis, we are focusing on the diagnosis of abnormalities in the main parts of female reproductive system: first, ovarian tumour diagnosis based on 2D B-mode ultrasound images of ovarian, and second, miscarriage identification based on 2D B- mode ultrasound image of the gestational sac. For this reason, we need to know about the abnormalities of ovarian tumours and gestational sac.

2.3 Ovarian Cancer

2.3.1 Overview of Ovarian Cancer

In general, the growth of any cell in the body is controlled by a certain gene. If this gene misses its function, it might lead to a formation of a tumour. A tumour that is not cancer is known as benign while a cancerous tumour is known as malignant. Benign tumours do not infect the tissues around them and nor spread to other parts of the body. However, malignant tumours have the ability to attack and destroy the tissues around them. Cancer cells can spread via the bloodstream or the lymph system and reach to other parts of the body (Tan, Agarwal et al. 2006). Ovarian cancer is strongly related to the age; although it could develop at any age, it is most likely to occur in older women especially from age 55 and over (Cancer Research UK 2014) Examples of benign ovarian tumours are teratoma, Endometrioma, Serous cystadenoma and Mucinous cystadenoma. Examples of malignant ovarian tumours are serous adenocarcinomas and mucinous adenocarcinomas, which originated from epithelial cells. According to (Bell 1991) epithelial tumours occur in pre-menopausal and post-menopausal age and 85% of ovarian cancer cases are epithelial cell (Chen, Ruiz et al. 2003). Although extensive research has been carried out, the real causes of ovarian cancer are still unknown. However, the two most effective factors linked to the risk of developing ovarian cancer are age and the presence of certain gene mutations. Other factors that may affect the risks of increasing ovarian cancer are described below:

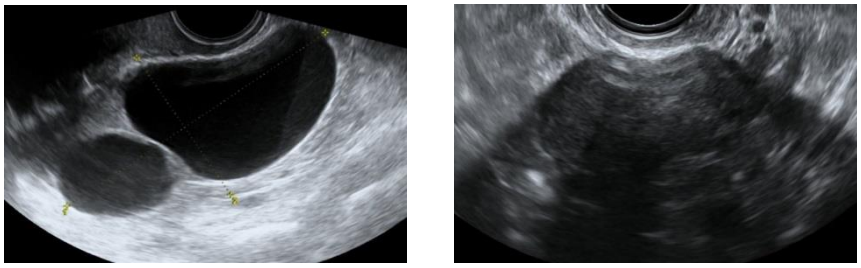
- Infertility, the risk is lower in women who had birth, compared to women who did not have children (Permuth-Wey and Sellers 2009).
- Family and previous cancers history, women with a mother or sister diagnosed with ovarian cancer have a higher risk of developing ovarian cancer. Similarly, women with a previous breast cancer have double the risk of ovarian cancer (Permuth-Wey and Sellers 2009).
- Miscarriage, multiple miscarriages are associated with an increased risk of ovarian cancer (Braem, Onland-Moret et al. 2012).

2.3.2 Ovarian Pathology

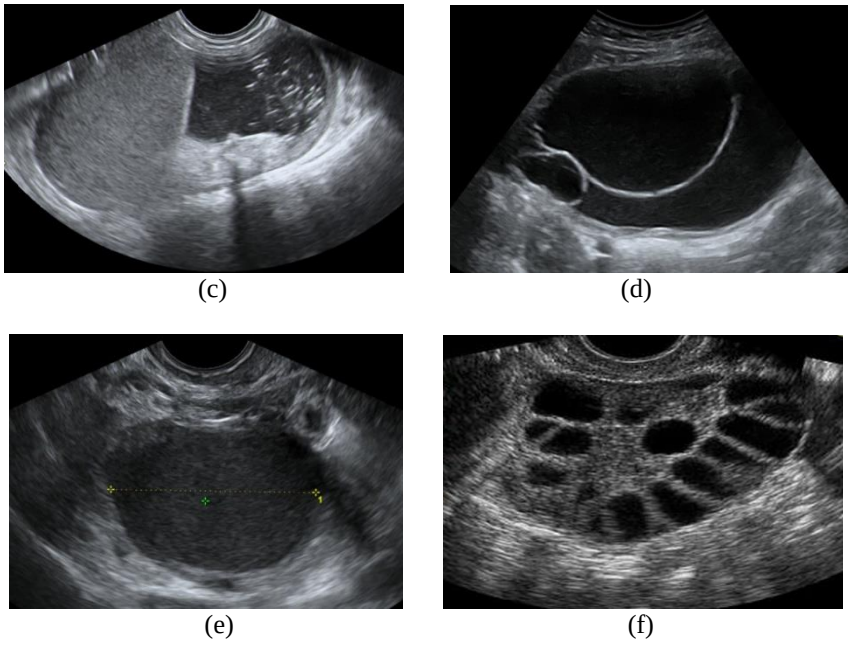
In this section, we describe the two main types of ovarian tumours benign and cancerous:

2.3.2.1 Physiological Cysts (Benign)

The term ovarian cyst often refers to a sac that has a fluid-filled develops in women of all ages. There are several different types of ovarian cysts. The most common one is known as a functional cyst, forms during the menstrual cycle and can be diagnosed by scanning (RL, Bogumil et al. 1971). Ovarian follicles larger than two centimetres in size are also considered as ovarian cysts (Loue and Sajatovic 2004). Beside ovarian cysts that are entirely filled with fluid only, other types of benign tumours additionally contain teeth, hair, fat, or blood clots inside the tumour. Different types of benign tumours may have different anatomy structures, and its correct diagnosis is important because each type of benign tumour or cystic needs particular treatment and care (RL, Bogumil et al. 1971). There are some cases of benign tumours that are still challenging to diagnose even by experts. Figure 2.14 illustrates some examples.



(a) (b)



(c) (d)

(e) (f)

Figure 2. 14: Examples of US images of different types of benign pathology (a) Simple cyst (b) Fibroma (c) Teratoma (d) Mucinous cystadenoma (e) Endometrioma (f) Polycystic ovaries

2.3.2.2 Malignant Ovarian Pathology (Cancer)

Malignant tumour or ovarian cancer starts and is caused by ovarian cells growing out of control. There are many types of ovarian cancer with their characteristics and behaviour. Ovarian cancer is generally grouped into three major types according to the tissue/cells that formed the lump: (i) epithelial, (ii) germ cell, and (iii) sex cord-stromal cell (Levine, Brown et al. 2010, Stany, Maxwell et al. 2010, UK 2014), Figure 2.15 shows clearly each cell.

1. **Epithelial tumours**. Approximately 85% of reported cases of ovarian cancer are epithelial. Epithelial ovarian tumours develop from the cells that cover the outer surface of the ovary (Chen, Ruiz et al. 2003).
2. **Germ cell tumours**. Such tumours grow from the egg-making cells in the ovary (Chen, Ruiz et al. 2003), and forms 5% - 10% of the reported cases of the cancerous ovarian tumour.
3. **Stromal cell tumours**. This type of ovarian cancer accounts for about 5% malignancy types. It develops from connective tissue cells that hold the ovary together and produce female hormones (Chen, Ruiz et al. 2003).

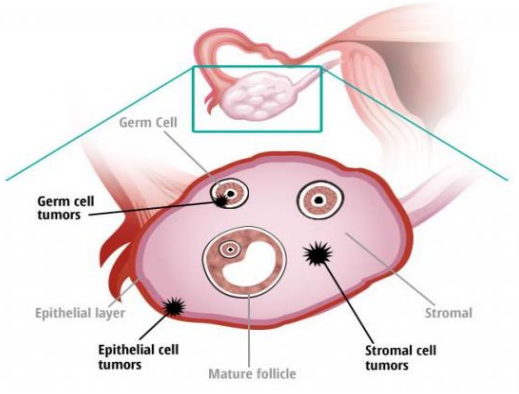


Figure 2. 15: Illustrating pathogenesis of ovarian cancer cell [45]

2.3.4 Stages of Ovarian Cancer

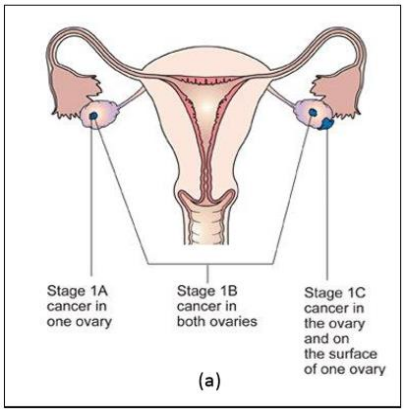
Ovarian cancer is divided into four stages. Stage 1 is limited to the ovary or ovaries while Stage 4 is wide-spread. Each stage is determined by the location of the cancer cells (Kurtz, Tsimikas et al. 1999, UK 2014). Figure 2.16 illustrate all the four stages of ovarian cancer.

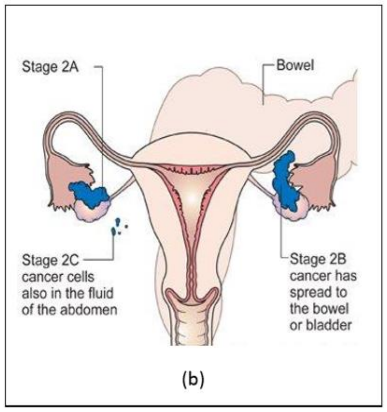
Stage 1: Cancer cells are found only in the ovaries (one or both), known as early cancer.

Stage 2: Cancer cells are found in one or both ovaries or in fallopian tubes and may have spread to other organs (such as the uterus, fallopian, etc.), within the pelvis.

Stage 3: Cancer cells have spread outside of the pelvic area to the abdomen. Tumours are very larger.

Stage 4: Cancer cells have spread to distant organs, such as the liver or lungs. This cancer is still known as ovarian cancer because the cancerous cells started in the ovaries.





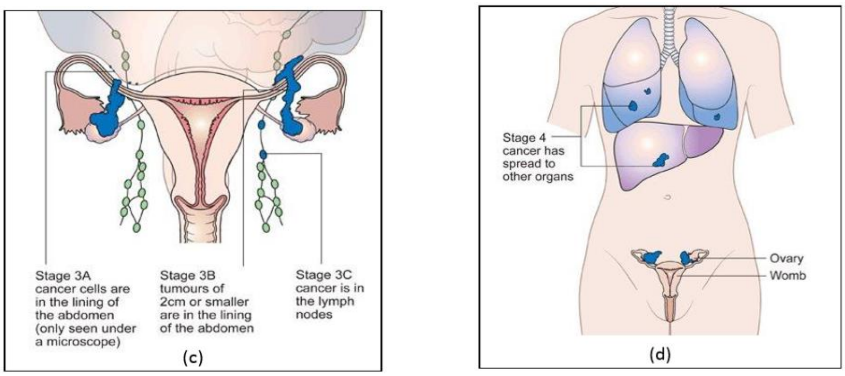


Figure 2. 16: Illustration of the extent of the spread of ovarian cancer - (a) - (d) for stages 1 to 4 (UK 2014)

Furthermore, there is another kind of tumour known as *Metastatic tumour* that started somewhere else (uterus, breast, pancreas, gastric, lung cancers) outside the ovaries but then spread to various body organs and arrived to the ovaries.

2.3.5 Diagnosis of Ovarian Cancer

Diagnosing ovarian cancer can be very challenging and the treatment entirely depends on the stage of the tumour (ACS 2014). Ovarian cancer is often diagnosed at a very late stage in most patients because there are no specific signs and/or clear symptoms of this disease. Diagnosis of ovarian cancer may be made using the following methods: physical examination, the blood test for CA-125 marker, and imaging procedure (Ultrasound / CT / MRI). The only definite ways to confirm exactly the diagnosing of ovarian cancer is by removing tissue from the suspicious area and examine it under a microscope (ACS 2014), which is known as a biopsy. We will now discuss the non-physical examination methods next.

2.3.5.1 Tumour Markers

Cancer Antigen 125 (CA-125) test identifies an abnormal level of a particular marker in the serum. It is used as a tumour marker to identify some types of cancer and is considered as an ideal for diagnosing ovarian cancer (Chen, Ruiz et al. 2003, Luo, Katsaros et al. 2003). Most often, the CA-125 test is used to assess the cancer response to treatment. The CA-125 marker is detected in approximately 80% of ovarian cancer, but not in early stages according to the Guideline No: 34 of the Royal College of Obstetricians and Gynaecologists (RCOG) (RCOG 2003). It remains the most useful tumour marker clinically that has been recommended by RCOG. CA-125 can be used together with B-mode transvaginal US for the assessment of ovarian cysts.

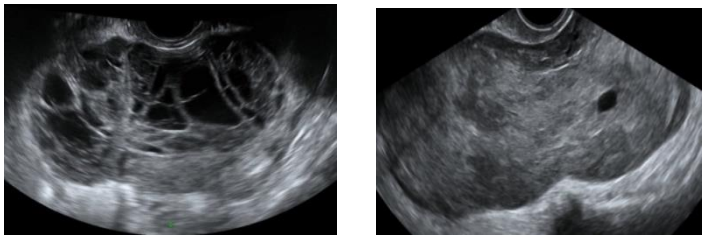
Research has started to be conducted to identify new biomarkers that can be used together with, or instead of, CA-125. Recently, Human epididymis secretory protein-4 (HE4) was found to be complementary to CA 125 for the detection of malignant disease. It has recently been reported that combining these two biomarkers increased overall sensitivity and specificity comparing with the use of a single biomarker (Moore, Brown et al. 2008).

2.3.5.2 Ultrasound-based Ovarian Cancer Diagnosis

Ultrasound-based assessment for ovarian tumours characterisations can yield two types of information depending on the type of US imaging modalities used: (i) morphological information which is based on B-mode images, and (ii) blood flow information which is based on Doppler imaging. Different levels of performance have been reported in the literature for various assessment schemes and models. In a study by (Kinkel, Hricak et al. 2000) it was found that the accuracy of ultrasound examiners using their subjective impression in reading real-time ultrasound images is 94% for B-mode ultrasound and 88% for colour Doppler flow. The only study that used both static B-mode and Doppler images together, has reported an accuracy of 85%. Ultrasound techniques that combine gray-scale ultrasound morphologic assessments with tumour vascularity imaging information (colour Doppler imaging) in a diagnostic system, are significantly better in ovarian lesion characterisation compared to using colour Doppler flow imaging, or gray-scale ultrasound morphologic alone (Kinkel, Hricak et al. 2000, Togashi 2003). Below we discuss the role of US imaging and morphology information in diagnosing ovarian tumours.

1. **B-mode Ultrasound Image in Diagnosing Ovarian Tumours**

Diagnosing complex ovarian cancer from B-mode ultrasound scan images is based on extracting morphological features which include the cystic and solid tumour structure, the presence and type of septations, and papillarities (Chen, Ruiz et al. 2003). Determining large size tumour is also another interesting features of ovarian cancer (Twickler and Moschos 2010). Figure 2.17 (a) and (b) are examples of B-mode images of an ovarian tumour.



(a) (b)

Figure 2. 17: Examples of ovarian large size tumours (a) Irregular multilocular tumour of largest diameter (b) Irregular solid tumour/Carcinosarcoma

While assessment of ultrasound findings is the best method for discriminating between benign and malignant tumours (Ameye, Valentin et al. 2009), this entirely depends on the observer experience (Wang, Itoh et al. 2002, Huang, Chen et al. 2008, Rocha, Campilho et al. 2011, Shung 2015). Standardising the terms and procedures in the interpretation of gynaecological ultrasound images is necessary for such diagnosis (Timmerman 2000). The 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"* (Rocha, Campilho et al. 2011). Several mathematical models and scoring systems for morphological features have been developed for discriminating between benign and malignant to increase the test performance. So far, several versions of the risk of malignancy index have been published. Examples of these scoring system are: Risk of Malignancy RMI, Logistic Regression models: LR1, LR2, Simple rules and ADNEX model. A study by (Chan and Selman 2006) has found that the most popular version of the scoring system used in the United Kingdom is the system that is published by the Royal College of Obstetricians and Gynaecologists in October 2003 (RCOG 2003).

a. Risk of Malignancy Index (RMI):

The risk of malignancy index (RMI) is a standardised index commonly used for clinical evaluation of patients with ovarian tumours that who are likely to have ovarian cancer. RMI is a very common type and it is used by many European countries (Vaes, Manchanda et al. 2012), and is defined by the formula:

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

- The score U is based on adding 1 point for each of the following five ultrasound image characteristics: Multilocular cyst, Evidence of solid areas, Evidence of

metastases, Presence of ascites, and Bilateral lesions. If the total ultrasound score is in the range 2–5, then U is given the value 3; otherwise, 1.

- M = 3 for all postmenopausal women, M=1 for all premenopausal women.
- CA-125 refers to the serum CA-125 measurement in u/ml and can vary between 0 and hundreds or even thousands of units. The normal range equal and less than 35u/ml.

The RCOG Guideline No.34 categorises the associated risk with different RMI ranges as follows:

Risk	**RMI**
Low	< 25
Moderate	25-255
High	>255

b. Logistic Regression Models :LR1 and LR2

The international IOTA organisation developed two models to estimate the likelihood of malignancy. The data was collected and analysed at nine centres which led to the development of two logistic regression models called LR1, LR2 (Timmerman, Testa et al. 2005, Timmerman, Van Calster et al. 2010). There are number of variables that have used in the scoring system, see Tables 2.1 1nd 2.2:

Table 2. 1: Logistic Regression models LR1

1	Personal history of ovarian cancer (yes = 1, no = 0)
2	Current hormonal therapy (yes = 1, no = 0)
3	Age of the patient (in years)
4	Maximum diameter of the lesion (in millimetres)
5	The presence of pain during the examination (yes = 1, no = 0)
6	The presence of ascites (yes = 1, no = 0)
7	The presence of blood flow within a solid papillary projection (yes = 1, no = 0)
8	Maximal diameter of the solid component
9	Irregular internal cyst walls (yes = 1, no = 0)
10	The presence of acoustic shadows (yes = 1, no = 0)
11	The colour score (1, 2, 3, or 4)
12	Presence of a purely solid tumour (yes = 1, no = 0)

The above table lists all 12 variables that are included in the model LR1. The LR1 model estimates the probability of malignancy. Between these 12 above variables, the pain and acoustic shadows tend to increase the possibility of having benign cases.

The model's estimated probability of malignancy equals to $1/(1 + e^{-z})$, where $z = -6.7468 + 0.0326(1) + 1.5513(2) + 1.1737(3) + 0.0496(4) + 1.1421(5) - 2.35\ 50(6) + 1.5985(7) - 0.9983(8) + 0.00841(9) - 0.8577\ (10) + 0.9281(11) + 0.4916(12)$, and *e* is the mathematical constant and base value of natural logarithms (Timmerman, Van Calster et al. 2010).

A simpler version using six variables was also developed based on the LR1 they have selected number of variables:

Table 2. 2: Logistic Regression models LR2

1	Age of the patient (in years)
2	The presence of ascites (yes = 1, no = 0)
3	The presence of blood flow within a solid papillary projection (yes = 1, no = 0)
4	Maximal diameter of the solid component expressed in millimetres
5	Irregular internal cyst walls (yes = 1, no = 0)
6	The presence of acoustic shadows (yes = 1, no = 0)

The model's estimated probability of malignancy of LR2 is equals to $1/(1 + e^{-z})$, where, $z = -5.3718 + 0.0354(1) + 1.6159(2) + 1.1768(3) + 0.0697(4) + 0.9586(5) - 2.9486(6)$ (Timmerman, Van Calster et al. 2010).

Both models have excellent diagnostic performance with a very high accuracy. Accuracy of LR1 and LR2 were found to be almost similar, and both LR1 and LR2 have a higher detection rate than RMI (Timmerman, Testa et al. 2005). Since LR2 requires fewer variables than LR1 does, LR2 is preferred.

c. Simple Rules

The multicentre IOTA has also recommended standardised examination techniques and standardised terms and definitions for examining patients and collecting data using ultrasound and Doppler ultrasound parameters to develop the newest rule. They have developed new rules to predict malignancy (M-rules) and (B-rules) to predict Benign. They chose various ultrasound variables or the combination of ultrasound variables that had the

highest positive predictive value with regard to malignancy. The variables are presented in Table 2.3 (Timmerman, Testa et al. 2008):

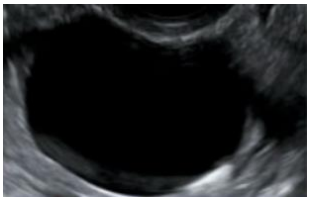
Table 2. 3: Ultrasound variables for Simple rules

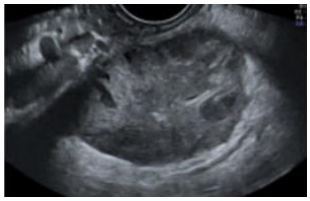
Rule No.	M-rules	B-rules
1	Irregular solid tumour	Unilocular cyst
2	Very high colour content on colour Doppler examination	Presence of solid components for which the largest solid component is <7 mm in largest diameter
3	Irregular multilocular - solid tumour with a largest diameter of at least 100mm	The presence of acoustic shadows
4	At least four papillary structures	Smooth multilocular tumour
5	The presence of ascites ascites	No detectable blood flow on Doppler examination

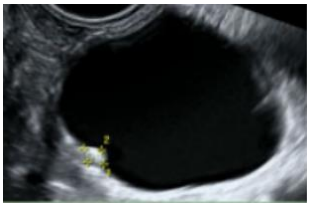
Combining the (M-rules) and (B-rules), IOTA has outlined three main malignancy prediction rules, as listed below in Table 2.4:

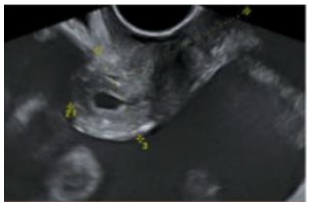
Table 2. 4: Three main rule for the that used for Simple rule
Rule 1:
If one or more *M features* were present in the absence of a *B feature*, they classified the tumour as ***Malignant***.
Rule 2:
If one or more *B features* were present in the absence of an *M feature*, we classified the tumour as ***Benign***.
Rule 3:
If both M features and B features were present or if none of the features was present, the simple rules were ***Inconclusive***.

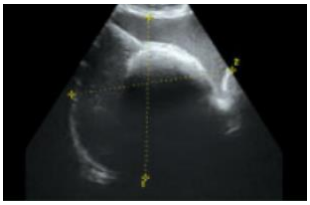
The diagnostic performance of the simple rule was shown to be relatively high and is used in many centres due to the easiness in applying these rules. Figure 2.18 shows some example ultrasound images with the ultrasound features that taken from (Kaijser, Bourne et al. 2013). The US images from B1–B5 display benign features; while the US images from M1–M5 show malignant features. The B5 and M5 are colouring US images, showing the status of blood flow.

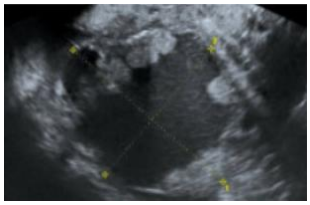

B1: Unilocular cyst

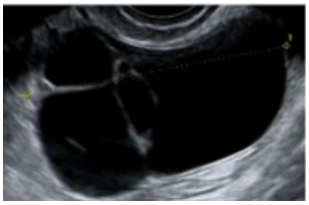

M1: Irregular soild tumour

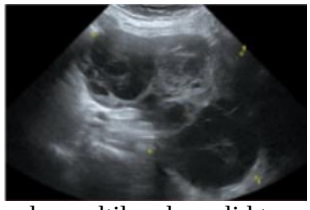

B2: Presence of solid components, with largest diameter < 7 mm

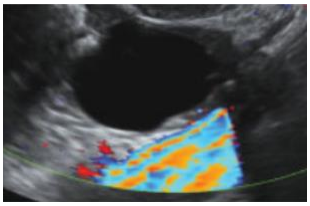

M2: Presence of ascites


B3: Presence of acoustic shadows


M3: At least four papillary structures


B4: Smooth multilocular tumour, with largest diameter < 100


M4: Irregular multilocular solid tumour, with largest diameter ≥ 100 mm


B5: No blood flow (colour score 1)

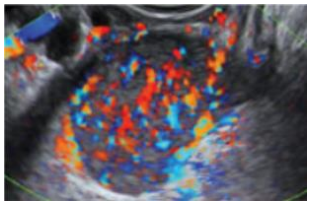

M5:Very strong blood flow (colour score 4)

Figure 2. 18: Ultrasound images and features used for Simple Rules

Table 2.5 sumarises the diagnosis accuracy rates together with sensitivity and specificity rates for 3 different models.

Table 2.5: Comparison performance of LR2, Simple rule and RMI derived by the IOTA group

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)
LR2	89	92	86
Simple rule	91.5	90	93
RMI	81	67	95

From this table, it is clear that the IOTA Simple rules have the best performance among the 3 models.

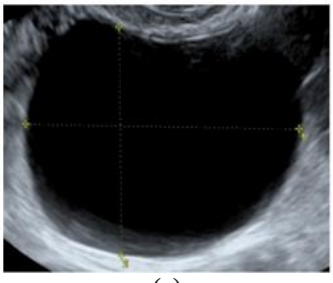
d. The ADNEX Model

Recently, the IOTA group published the Assessment of Different NEoplasias in the adneXa (ADNEX) model, the model is considered as a first risk model that distinguishes between benign and four types of malignant ovarian tumours: borderline tumours, stage I cancer, stage II-IV cancer, and secondary metastatic cancer. This approach is novel compared to the existing models (RMI, LR1, LR2, simple rule) that only differentiate between benign and malignant tumours (Van Calster, Van Hoorde et al. 2015).

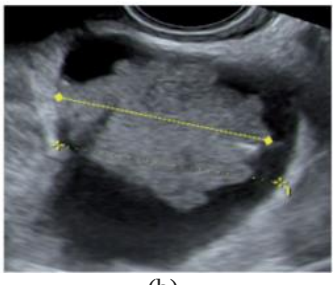
The ADNEX model consists of three clinical predictors and six ultrasound predictors. The clinical predictors are age (in years), serum CA-125 (U/ml) and type of center to which the patient has referred for ultrasound examination. Type of centre has been divided into oncology centers versus other hospitals (Van Calster, Van Hoorde et al. 2015). Table 2.6 below shows clearly the six variables that are used in the ADNEX model, Figure 2.19 shows examples of the ultrasound image of the ovarian tumour with an explanation about the six variable of ADNEX model.

Table 2. 6: Six ultrasound predictors for ADNEX model

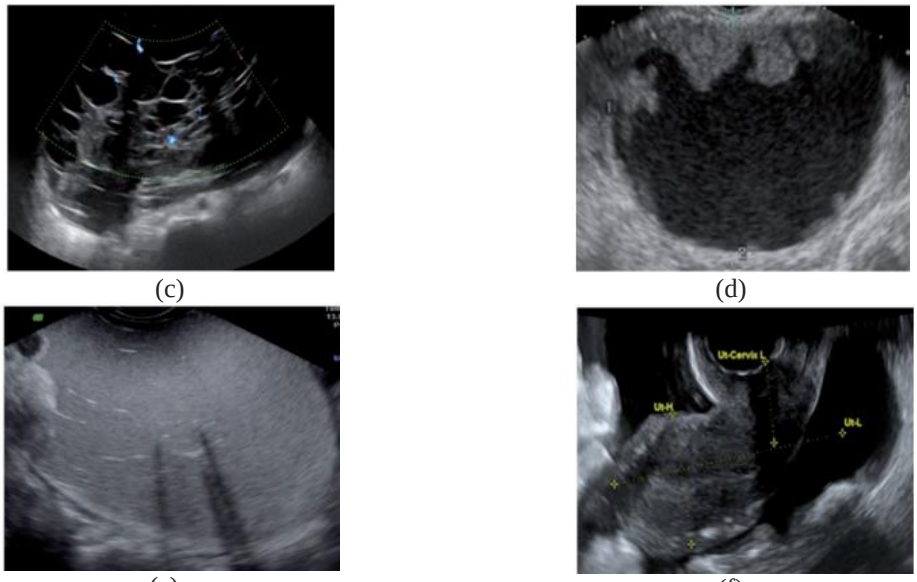
1	Maximal diameter of the lesion (mm)
2	Proportion of solid tissue (%)
3	Number of papillary projections (0, 1, 2, 3, > 3)
4	Presence of more than 10 cyst locules (yes = 1, no = 0)
5	Acoustic shadows (yes = 1, no = 0)
6	Presence of ascites (yes = 1, no = 0)



(a)



(b)



(c) (d)

(e) (f)

Figure 2. 19: Ultrasound image characteristics chosen as predictors for ADNEX model (Van Calster, Van Hoorde et al. 2015) (a) Maximal diameter of the lesion (mm) (b) Proportion of solid tissue (%) (c)More than 10 cyst locules (d)Number of papillary projections (0, 1, 2, 3, morethan 3) (e)Acoustic shadows (yes = 1, no = 0) (f)Presence of ascites (yes = 1, no = 0)

Table 2.7 presents the risk rates for the different final diagnoses using the combined data from IOTA phase 1-3 dataset (n = 5909). More details can be found in (Van Calster, Van Hoorde et al. 2015).

Table 2. 7: Baseline risk for the different final diagnoses using ADNEX model (Van Calster, Van Hoorde et al. 2015)

	Overall (%)	Oncology centre (%)	Other centre (%)
Benign	68.2	48.8	83.7
Malignant	31.8	51.2	16.3
Borderline	6.3	9.3	3.8
Stage I invasive	7.4	10.3	4.6
Stage II-IV cancer	14.1	24.3	6.4
Secondarymetastatic	4.0%	7.3%	1.6%

Although the use of a scoring system helps to improve the test performance, according to (Gramellini, Fieni et al. 2008), the existence of several scoring systems may cause difficulties and inconsistency in clinical practices generally for the following reasons:

(i) Experienced ultrasound examiners use additional demographic information when they are estimating the tumour types, the less experienced ultrasound examiner have difficulty in acquiring the ultrasound morphology information that needs experience and training.

(ii) A high percentage of tumours do not conform to these rules. Not all masses clearly yield the relevant information.

(iii) The scoring systems work well with tumours that are easily classifiable using pattern recognition, but less well with tumours that are difficult to classify using pattern recognition (Timmerman, Testa et al. 2008). In addition, there is no general agreement about some morphology information while trying to transfer this information to a scoring system.

2. Doppler Ultrasound Images

The Doppler ultrasound imaging system provides additional technology to measure the blood flow (Hoskins, Martin et al. 2010). Colour Doppler imaging technology is the efficient method and first step to detect the malignant lesion. This feature allows the assessment of tumour vascularity because malignant masses are expected to have the active blood vessel in comparison to normal masses. The blood flow detection within a mass is an indication of malignancy (Kinkel, Hricak et al. 2000). It has mainly been used for distinguishing between benign and malignant lesions (Alcazar 2006). Assigning a colour score to the amount of blood flow inside the tumour is dependent on the subjective assessment of the ultrasound examiner see Table 2.8.

Table 2. 8: Colour score based on Doppler ultrasound image

Colour score	Amount of blood flow based on color Doppler image
1	No blood flow
2	Minimal blood flow
3	Moderate blood flow
4	High blood flow

According to (Timmerman 2000) using colour Doppler ultrasound is very critical because: the assessment is (i) subjective, and (ii) it depends on the quality of the equipment and the settings.

3. 3D Ultrasound Imaging for Ovarian Tumour Assessment

The use of 3D scan imaging in the assessment of tumour may add some extra information that is not available from standard 2D scanning in ovarian cancer diagnosis. However, a comparison study of 2D and 3D power Doppler imaging diagnostic capabilities for the

prediction of ovarian cancer has concluded that 3D power Doppler imaging did not have a better diagnostic performance over 2D power-doppler imaging for the discrimination between benign and malignant tumours, (Alcazar 2006). Further studies have conducted by a number of research groups (e.g. (Kurjak and Kupesic 2003, Testa, Ajossa et al. 2005, Jokubkiene, Sladkevicius et al. 2007, Dai, Hata et al. 2008), they concluded that 3D US images is effects on the characterising of ovarian tumours and improve the diagnostic accuracy, but later in a study (Alcázar and Jurado 2011) concluded that further studies are needed to establish the role of 3D ultrasound in clinical practice in gynaecological oncology.

2.4 Early Pregnancy and Miscarriage

2.4.1 Pregnancy Overview

Throughout the fertile life time of a woman, ovulation should occur 14 days before a woman's next menstrual period begins (Edwards and Steptoe 1975, Findlay, Kerr et al. 2009). An egg can be fertilised 12 - 24 hours after ovulation

Within 24-hours after fertilisation, the egg begins dividing rapidly into many cells. The fertilised egg, called a zygote, continues to divide (then called a blastocyst) as it passes slowly through the fallopian tube to the uterus. It next attaches itself to the uterus in a process called Implantation (Edwards and Steptoe 1975, Brown 2014). Figure 2.18 presents all stages from Ovulation to Implantation.

Within three weeks, while the blastocyst cells continue to grow, and the endometrium develops around it, a structure is formed known as the Gestation Sac (GS). The GS is the first sign and measurable element of an early pregnancy by ultrasonography.

Baby developing is called an embryo from the moment of conception to the eight week of pregnancy. After the eighth weeks and until the moment of birth, the developing baby is called a fetus (Edwards and Steptoe 1975, Campion, Doubilet et al. 2013). Figure 2.20 clearly show the stages of Ovulation and fertilisation.

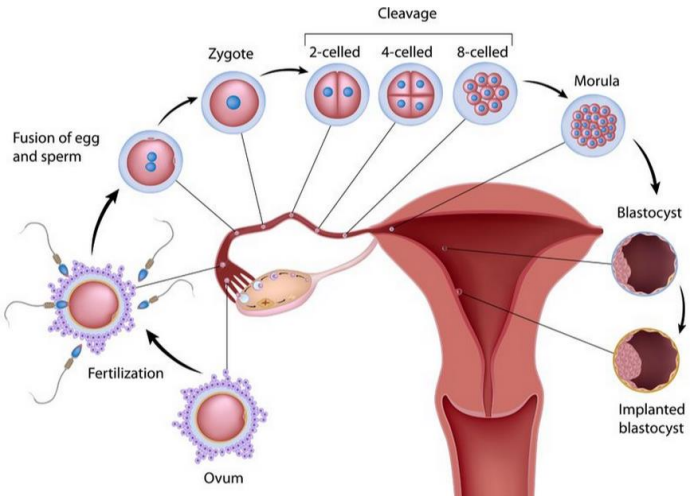


Figure 2. 20: Female Ovary Showing Ovulation and fertilization (Campion, Doubilet et al. 2013)

A normal pregnancy is 40 weeks long (plus or minus two weeks). The first symptoms of pregnancy are missed the menstrual period. Pregnancy tests, usually performed on the urine, are designed to demonstrate the presence of Human Chorionic Gonadotropin (hCG) (Campion, Doubilet et al. 2013). If the urine test is positive, then the first scan will be offered to find out the age of the gestational sac and determine the due date. Figure 2.21 clearly present an example of the anatomical structure of the early pregnancy. The anatomical structures of the early pregnancy are characterised by:

1. Amniotic sac. It is a bag of clear, pale fluid inside the uterus where the fetus starts to develop and grow. It is connected to the yolk sac at the later stage (Omaha Nebraska 2013).
2. Gestational sac (GS). It is a structure that surrounds an embryo, in the very early stages of pregnancy. This sac grows by approximately 1mm per day. Figure 2.22 (a) show ultrasound image of the gestational sac.
3. Yolk sac. It is the ring-shaped structure identified within the gestational sac. Figure 2.22 (b) shows YS inside the GS. The yolk sac will be visible before a clearly definable “embryonic pole” (Omaha Nebraska 2013).
4. Heartbeat. It will be seen along with the yolk sac. Failure to identify fetal heartbeat is a sign of abnormal pregnancy this may lead to miscarriage later (Omaha Nebraska 2013). Figure 2.22 (c) show the embryo attached to the YS inside the GS.

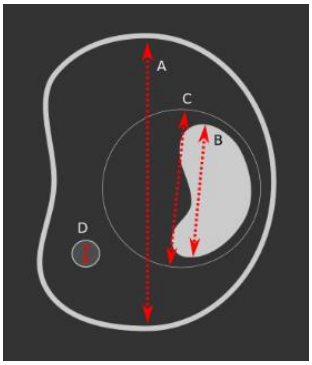
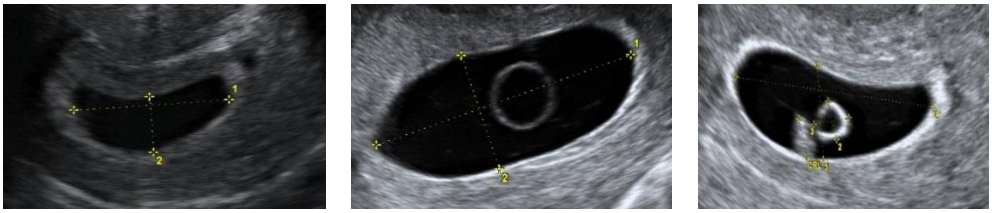


Figure 2. 21: The anatomical structures of the early pregnancy A: Gestational sac (GS), B: Crown rump length (CRL) of embryo, C: Amniotic sac and D: Yolk sac



(a) (b) (c)

Figure 2. 22: Examples shows the ultrasound images of a very beginning of pregnancy until developing the embryo (a) Gestational Sac (b) YS within GS (c) embryo attached with YS within GS

2.4.2 Early Ultrasound Scan

A positive pregnancy test will normally be followed by a first scan on the 12th week pregnancy. However, if the pregnant woman feels abdominal pain or bleeding and any abnormalities then she will be sent immediately to an early pregnancy scan unit to:

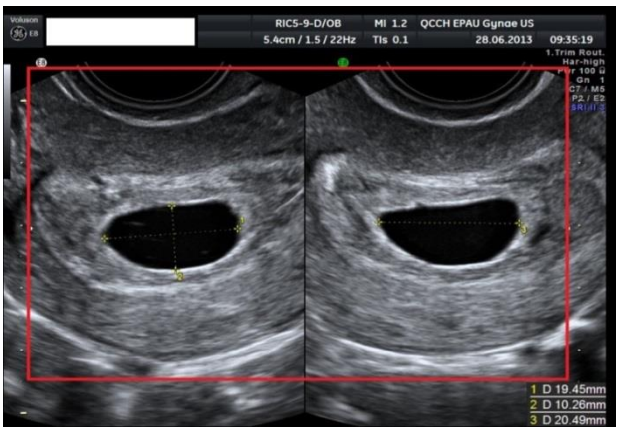
- Measure the gestational sac,
- Monitor the development of the yolk sac YS within gestational sac GS,
- Detect the embryo, and
- Detect the heartbeat of the embryo.

2.4.3 Miscarriage Diagnosis

Miscarriage happens when a woman loses her pregnancy in the first three months. This means the gestational sac stops growing or the embryo stop developing (Jurkovic, Overton et al. 2013, ACOG 2015). Most miscarriages are caused by a "chromosomal fault within the cells of the developing pregnancy". For an Embryo to grow properly must have the right number of chromosomes. In very rare cases the uterus shape or the mother's immune system can cause miscarriage as well (RCOG 2015).

The first measurable sign of an early pregnancy are the geometric characteristics of the Gestational Sac (GS). The first feature calculated is the mean sac diameter MSD of the GS.

MSD is the average of the three diameters that taken from two planes of the GS. These two planes are sagittal and transverse planes. Figure 2.23 presents an example image of the GS taken in respective sagittal and transverse planes. The red rectangle represents the main fan area. The GS is the dark region in the centre. The area outside the red rectangle is called the margin area. It is used to display information about the patient (blocked for anonymity), the date and time that the image was taken and the ultrasound machine setting. The Figure shows the three manual measurements of the GS size marked by yellow dotted lines. The measurement results in millimetres are present in the margin area. There are other signs of likely miscarriage. For example, the border of the GS appears irregular in its shape. The same process of scan is repeated at the later stage when the yolk sac is starting grow to check the pregnancy. Figure 2.24 is the second stages when the YS is growing inside GS.



Sagittal plane view Transverse plane view

Figure 2. 23: Ultrasound image of GS in Sagittal and Transverse planes

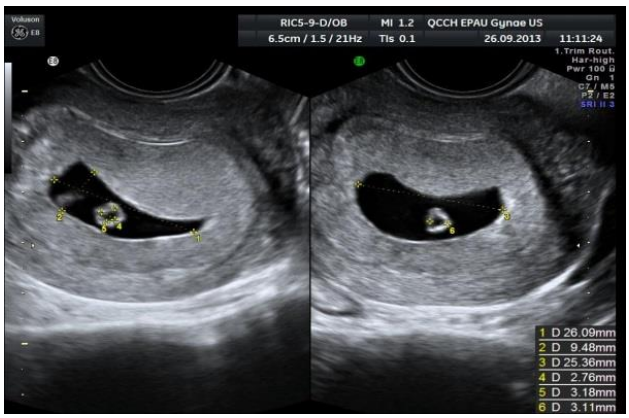


Figure 2. 24: Show later stage of pregnancy when YS is grows inside the GS

In the later stage when the GS and YS are developed without any problem, the embryo starts to grow and attaches to the YS. Measuring the embryo is quite different from the sacs. Crown Rump Length (CRL) is used to measure the length of embryos and from the top of the head

(crown) to the bottom of the buttocks (rump) (Omaha Nebraska 2013). After that, attention focuses on detecting the embryo heartbeat by switching to the colour Doppler ultrasound image. The most important aspects of normal pregnancy in ultrasound image scanning are (Campion, Doubilet et al. 2013):

- The Gestational Sac GS should be visible in 4.5 -5.5 weeks.
- The yolk sac is a definite within 5.5 - 6 weeks.
- Embryo with heart beat within 6- 6.2 weeks,

2.4.3.1 Existing Rules for Diagnosing Early Pregnancy Failure

Analysis of ultrasound images can assist diagnosis of possible miscarriages at a very early stage in pregnancy i.e. in the first three months, which is medically called the first trimester (Geirsson and Busby-Earle 1991). Monitoring pregnant women at this stage enable the specialists to evaluate the development, growth, well-being of fetus and to predict the due date (Kaur and Kaur 2011). Different criteria have been used to diagnose miscarriage. For example, the American College of Radiology guideline defines miscarriage as being an empty GS with a Mean Sac Diameter (MSD) greater than or equal to 16 mm (Levi, Lyons et al. 1990, Bourne and Bottomley 2012) whereas the Royal College of Obstetricians and Gynaecologists in the United Kingdom recognises miscarriage as being an empty sac with MSD greater than or equal to 20 mm (NICE clinical guideline,154 2012). A recent study reviewed the rule in (NICE clinical guideline 154, 2012) and concluded that an empty GS with MSD greater than or equal to 25mm should be introduced as a new guideline to minimise the risk of false positive diagnosis of miscarriage (NICE clinical guideline,154 2012). Table 2.9 below represent the most recent cut-offs for miscarriage diagnosis.

Table 2. 9: Miscarriage identification cut-offs according to the NICE guidline 154, 2012

1	Mean gestational sac diameter (MSD) of ≥ 25 mm with no obvious Yolk sac (YS)
2	Mean sac diameter of (GS and YS) of ≥ 25 and no embryo define
3	Crown-rump length (CRL) of ≥7 mm and no heart beat detection

2.4.3.2 Accurate Prediction of Pregnancy Viability Based on Simple Scoring System

Physicians have an important role in accurate and certain miscarriage diagnosis. The exact incidence of miscarriage is difficult to assess, but in the above section, we referred to a number of expert guidelines on specific cut-offs of miscarriage diagnosis. In addition,

recently the group of researcher have developed a new simple scoring system based on an individual demographic and ultrasound factors, such as maternal age and vaginal bleeding (Bottomley, Van Belle et al. 2013). These factors, combined with ultrasound diagnosis based on Table 11 rules, have demonstrated that routine clinical information about the mother and her symptoms as well as some simple ultrasound parameters can be combined to very accurately determine viable pregnancy at the end of the first trimester. The combination of such factors (age, bleeding score, gestational age, mean GS and YS sizes and the presence of fetal cardiac activity) provide a more reliable prediction of viability than any of the individual factors alone. This model is easily transferable to the clinical setting by way of the presented scoring system, where both maternal and ultrasound variables are easily available to the physician and can be used to make an accurate assessment of the likelihood of viability. Table 2.10 present examples of the scoring system variables. More details about a scoring system are in (Bottomley, Van Belle et al. 2013).

Table 2. 10: Demographic and symptom variables

Variables	Categories	Points
Maternal age (years)	< 35 35-39 ≥ 40	0 -1 -2
Bleeding score	0 1 2 3 4	0 -1 -2 -3 -4

In addition, the following factors are also associated with miscarriage (Brown, Emerson et al. 1990)

- Uterine Abnormalities such as T-shaped uterus or low uterine position
- Irregular contour of GS
- Polycystic ovary syndrome
- Aged >40 years
- Smoking
- Diabetes
- Thyroid disease

2.5 Summary and Conclusion

This chapter reviewed the background on medical imaging and its use as a reliable tool that helps to diagnose, treat and monitor patients. The US is considered as an ideal imaging system for gynaecological diagnosis abnormalities. Ovarian cancer is one of the most common cancers in women and is the leading cause of death from gynaecologic malignancies. Miscarriage is a very common problem to happen during the first 12 weeks of pregnancy. Based on a recent study there is also a strong link between miscarriage and ovarian cancer development (Braem, Onland-Moret et al. 2012). This means that finding effective computer-based solutions to solve both problems is timely and desirable, and any positive contributions that can be made will bring benefit to improve patient care and reduce the loss of life.

The interpretation of the ultrasound image, however, is highly dependent on the ability and experience of the observer. Very often, an observer makes a diagnostic decision with a level of certainty. In certain cases, the observer may not be entirely certain in his/her diagnostic decisions, and often in those cases, expert opinions, sometimes from more than one expert, may be sought in assisting the diagnosis by combining the decisions into a final sensible outcome. This process will increase the level of confidence and lead to better diagnosis results. Also quite often, there will be different opinions from different experts in diagnosing a specific case, resulting in an "inclusive case" that needs to be further examined.

Regrettably, the limitations in the human eye-brain visual system, reader fatigue, distraction, and the presence of overlapping structures in images may cause detection and interpretation errors. This increases the number of false-positive and false negative results. Therefore, there is a need to develop a computer-aided diagnosis system CAD for the purpose of detecting the gynaecological abnormality in early stage. CAD may be used as a support tool for automating the measurements of the values for certain parameters in order to improve measurement precision and avoid intra and inter observer variations in manual measurements of the parameters. Besides, modern image processing techniques that have been developed and matured in the past few decades may offer alternative and effective features that can be directly extracted from US images that are "outside" of the known parameters to medical experts. Indeed, we have already started to witness rapid developments of such "alternative" descriptive features in various fields of application in recent years.

In the next chapter, we will present our research methodology and a road map of our investigation for automatic solutions for accurate classification of ovarian tumours and efficient prediction of miscarriage cases.

CHAPTER 3

RESEARCH FRAMEWORK FOR GYANCOLOGICAL US IMAGES ANALYSIS AND DIAGNOSIS

In Chapter 2, we presented background information on medical imaging system in diagnosing gynaecological abnormalities. We highlighted the role of B-mode ultrasound images in the diagnosis of the most common diseases in female reproductive organs with a focus on miscarriage cases and ovarian tumours. Moreover, we explained different types of ovarian tumours and different type of IOTA models for their diagnosis in the clinical setup. Furthermore, for miscarriage, we explained the details of the process that lead to miscarriage and the NICE guideline on the stage of miscarriage identification. We have also explained the manual discrimination between positive and negative cases in early miscarriage identification and ovarian tumour diagnosis from real-time ultrasound images. This thesis argues that the manual process involves multiple subjective decisions which may increase the inter- and intra-observer variations that could subsequently leads to difficulties and even errors in the diagnosis stage based on (Wang, Itoh et al. 2002, Giger, Chan et al. 2008). In fact, interpretation of ultrasound images is highly dependent on the medical expertise and experience of the observer and this is subject to considerable variation, see (Wang, Itoh et al. 2002, Huang, Chen et al. 2008, Shung 2015). Therefore, reliable automatic diagnosis can provide a decision support tool that helps experts to increase the accuracy of diagnosis.

This chapter outlines the framework of our work in the area of automatic analysis of B-mode ultrasound images of ovarian tumours and gestational sac for miscarriage identification. This framework consists of the main tasks of image processing, feature extraction, classification techniques, evaluation and experimental protocols for the gynaecological abnormalities under investigation. The chapter is organised as follows: Section 3.1 describes image processing and analysis stages used in typical image analysis systems. Section 3.2 explains our framework to identify different types of ovarian tumours. The framework of the early miscarriage identification from US images of Gestational sac is described in section 3.3. In section 3.4 we explained the datasets of US images used throughout this thesis. Section 3.5 describes different types of classifier methods that used in this study. A concept of fusion methodology briefly is explained in section 3.6. Experimental protocols and performance

evaluation of this thesis in ovarian tumour classification and miscarriage identification are described in section 3.7. Summary of this chapter is presented in Section 3.8.

3.1 Ultrasound Image Processing and Analysis

Digital images contain a huge quantity of data and information the interpretation of which needs effective image processing and analysis solutions. For automatic diagnostic purposes, effective image features that model the medically relevant information need to be first extracted efficiently and reliably. Reliability of the extracted features is dependent on the image quality which cannot be guaranteed at the capturing stage, and thus improving image quality is an essential task. Analysing the automatically extracted features is the main diagnostic objective in our research which will require the use of advanced machine learning techniques and classification tools. Computer-based analysis plays an important role in extracting useful information using automatic and semi-automatic algorithms to perform image analysis with high levels of accuracy and speed. Automatic image processing and analysis is desirable to help the sonographer/gynaecologist to get accurate results for better and reliable diagnosis.

In general, image processing encompasses a variety of application-dependent procedures that relate to the desired objectives. For our purposes, it can be defined as the manipulation of US images in preparation of extracting gynaecology-related data and information features from the images. This preparation often begins by quality enhancement tasks which may require an alternative representation of the image from the captured spatial domain representation.

Image processing techniques are based on image operators that transform any input image to produce an output image model, as described in equation (3.1), suitable for feature extraction and analysis. Image operators either manipulate the spatial information (transformation in the spatial domain) or manipulate the frequency information (transformation to the frequency domain) (Gonzalez, Woods et al. 2004).

$$g(x, y) = T(f(x, y)) \tag{3.1}$$

where g (x, y) is the output image, f(x, y) is the input and T is the image operator

The main steps of medical image processing and analysing are: image enhancement, de-noising, segmentation, feature extraction and classification. Figure 3.1 shows the general block diagram of the framework of automatic ultrasound image processing and analysis.

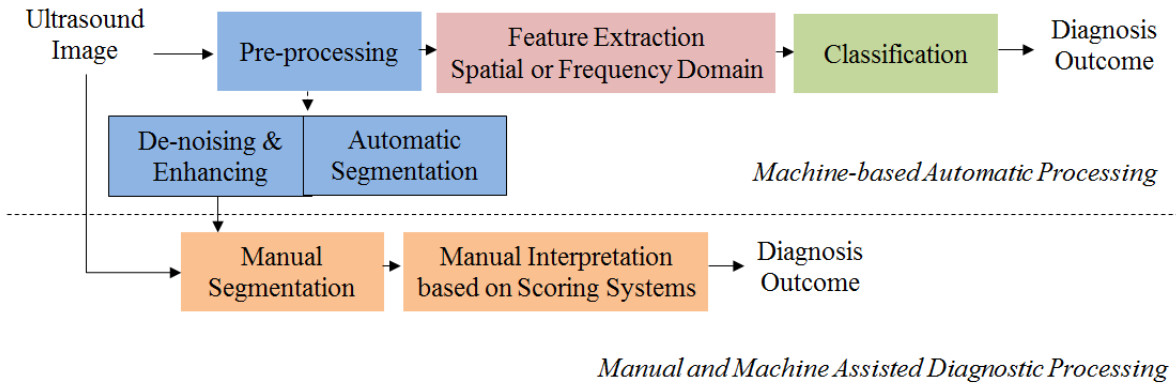


Figure 3. 1: Block diagram of the major steps of automatic vs manual ultrasound image analysis

1) Image Pre-processing: Ultrasound images are affected by noise such as speckle noise which we suppress using filtering techniques (de-speckling) to enhance the images for better visualisation and to highlight hidden information/texture.

2) Image Segmentation: This stage partitions the image into a number of small portions and separates the region of interest (ROI) from the background.

3) Feature Extraction: Features relevant to general US image analysis include measurement types of features on the ROI (e.g. diameters of a tumour or the GS) plus other more image-content related features in the ROI that distinguish it from its exterior and variation of which may characterise different classes under investigations (e.g. texture parameters in a tumour or the GS). In this work, the ROI digital features are either extracted from the US image spatial or frequency domain. Extracted features are the main input into classification procedures.

4) Classification: The extracted features are input into various classification methods with the aim of building a classification model (classifier) to label the ROI as Normal or Abnormal. A variety of classifiers are available in the literature that have been developed for general pattern recognition tasks and forms the essential component of machine learning.

3.2 Overview of the Proposed Ovarian Tumour Classification

The diagram in Figure 3.2 depicts the components of the various US image-based automatic steps outlined at the end of section 3.1, that lead to the diagnosis of ovarian tumour. The static B-mode ultrasound image of the ovarian tumours is the input.

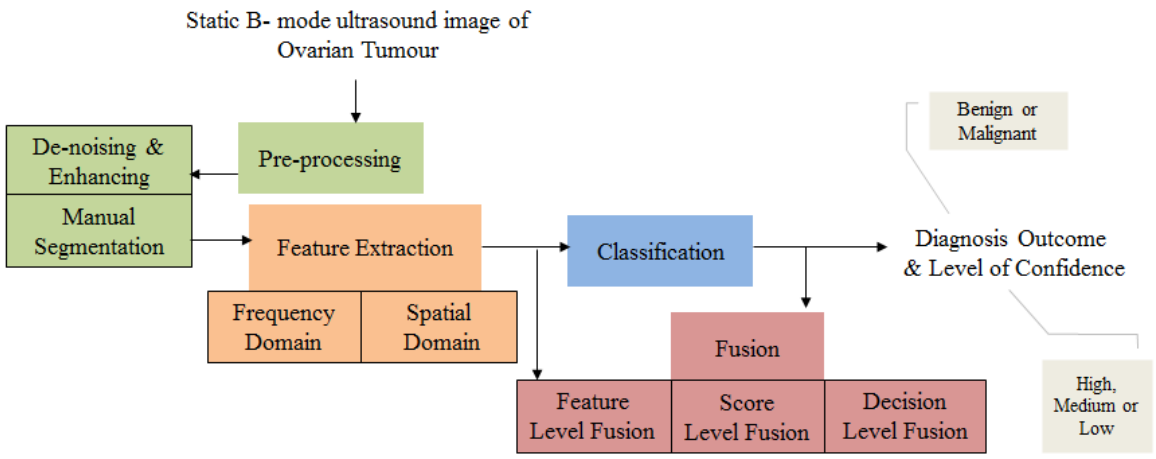


Figure 3. 2: Block diagram of the process of analysing B-mode ultrasound images of ovarian tumours

We shall now describe our research methodology to implement the various steps of the above process.

1. Usually ultrasound images are corrupted by random noise during data acquisition processes. For this reason, we need to apply image processing filters to remove the artefact and to enhance the images by highlighting the details of the region of interest ROI, i.e. the inner and outer border and/or the texture inside the ROI. Removing the noise from ultrasound images helps a human observer (radiologist) in clinics to easily and accurately detect the ROI and capture the anatomical information inside the region. Furthermore, removing noise and enhancing the images in the automatic analysing process will also lead to more accurate segmentation of the ROI, easier extraction of effective features, and eventually more accurate automatic diagnosis results. Our approach to deal with the issue of noise removal will benefit from existing knowledge about common noise removal techniques. We used the *Non-local means filter* that differs from traditional filters in that it attempts to exploit redundancy and self-similarity in the input image, and hence enables the removal of noise while keeping almost all meaningful information. More details are given in Chapter 4, Section 4.3.2.
2. After removing the noise, image will be ready for the segmentation of the ROI from the image background. Various traditional methods, such as Thresholding, Region growing, etc. are available for segmenting the ROI. It has been recognised in the literature that automatic segmentation of ultrasound images is a complex, challenging and domain-specific task, and still an open problem for research. In order for our research not to be bogged down at this single step of the automatic diagnosis process,

we adopted a semi-automatic method to deal with the very complex natures of the US images we encounter. More details about these complex natures and our method can be found in Chapter 4, Section 4.3.3.

3. The next stage is the extraction of the most important features from the segmented images of tumours. These features could be in spatial or frequency domain. The extracted features for automatic tumour diagnosis are additional and different from those features that extracted manually by the radiologist. Although the size/volume of the tumour is an important feature for the manual diagnosis by the experts, in our research we focus on texture-related features rather than the measurement features. These features can be easily saved in a vector and labelled by the class that belong to. This part of the research is described in details in Chapter 5.
4. The next obvious step, in our work, is to feed the features extracted from an ovarian US image to an appropriate classifier in order to produce a diagnostic score/decision. Although there are varieties of types of benign and malignant tumours, in this work we shall only focus on using binary classifiers for the two states of the tumours. The extracted features from any input US image will be fed into such a binary classifier for diagnosis as a benign or a malignant tumour. In Section 3.5, we briefly discuss the different classifiers employed in this work. The classification results based on extracted features are presents in Chapter 5.
5. The fact that, we have been extracting different feature vectors from different image representation domains means that we can get diagnosis results by different schemes that use different feature vectors and/or different classifiers. This common characteristic of many pattern recognition problems has been exploited to improve accuracy of recognition/classification by using the concept of *Fusion* at different levels. A main innovative feature of our research framework, for ovarian tumour classification, will be the investigation of various fusion schemes. In particular, our work will include fusion at the *feature level* prior to classification, at the *score level* post classification, and at the *decision level* also post-classification. This will be discussed in Chapter 6. Our novel decision level fusion will be described in Chapter 6, Section 6.5.3.
6. One of the main issues that have been often overlooked in the case general automatic pattern recognition and classification is how reliable is the decision output by the classifier. To evaluate false positive and false negative cases, the specificity and sensitivity rates are used to provide a measure of confidence in the accuracy rate of

the classification procedure. However, individual classification decisions are based on some thresholded scores, and surely the strength of such a decision must be dependent how near the actual score is to the set thresholds. Therefore, in our work we introduce a quantitative concept of decision strength into the classification stage which will be based on how far a sample score is from the cut-off threshold. We discuss this concept in details in Chapter 6, section 6.3 and we argue that the level of decision strength provide more insight into the real performance of the adopted classification methods and sets the level of confidence in the individual diagnosis.

3.3 Overview of the Proposed Early Miscarriage Diagnosis

The static B-mode ultrasound image of gestational sac GS is the starting input of our work in relation to miscarriage. Each image is presented in two planes (Sagittal and Transverse). Figure 3.3 describes our work in the area of miscarriage identification from B-mode ultrasound image of the gestational sac.

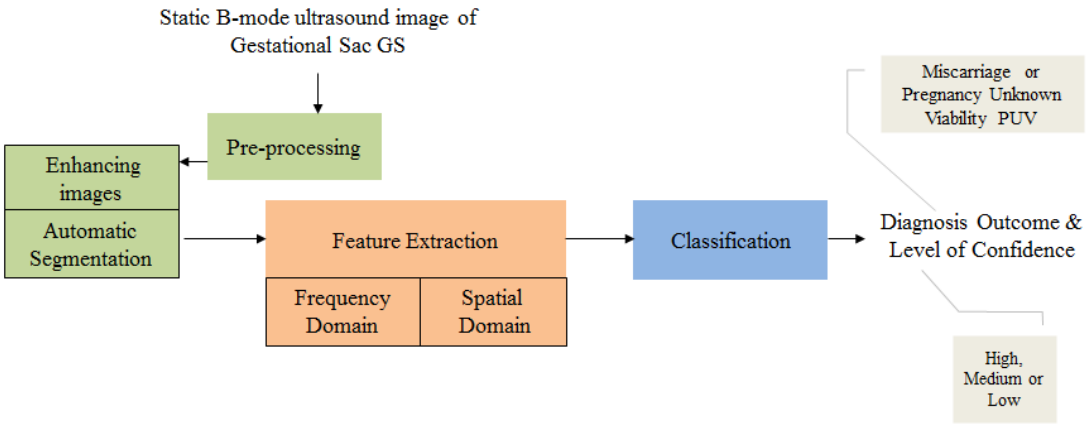


Figure 3. 3: The block diagram of the main process behind human B-mode ultrasound image of GS classification

The various steps of the developed scheme and the relevant research investigations conducted for this case differ from the steps of the ovarian tumour case only in that the main geometric features of the GS are known in the application domain and some additional geometric measurements of GS are also explored. Below is the summary of this process:

1. The input US image is enhanced to produce a more visibly highlighted GS.
2. Automatically segment the GS as the ROI and remove the background from the enhanced image.

3. Apply an automatic procedure to extract the three well-known diameter measurements of the ellipsoid shape GS and calculate their mean which is called *Mean Sac Diameter (MSD)*. In addition, we extracted a new set of image-based features such as (Perimeter, Volume, Area, etc.). More details discussed in Chapter 5, Section 5.3. All these features are saved in a feature vector.
4. These automatically extracted features are then used in the classification stage to identify the miscarriage cases from the pregnancy of unknown viability PUV cases.
5. Additionally, as in the above case, we also introduce the concept of level of confidence into the classification result. We argue that the level of decision strength provide more insight than other classical classification methods and makes the proposed decision closer to the practical diagnosis.

Each stage of our work for both areas will be explained in details in the followed Chapters.

3.4 Datasets used in this Study

Testing the performance of any automatic pattern recognition and medical diagnostic schemes rely heavily on the choice of the datasets of relevant US images that will be used in the experimental work using the appropriate evaluation protocols. In general, the classes of the ovarian tumours/GS are known in these sample images in order to be able to determine success and failure of the automatic classification scheme. It is essential that the sources of such datasets are widely known and recognised for credibility and relevance of their work.

3.4.1 Static B-mode Ultrasound Images of Ovarian Tumours

All static ultrasound images of ovarian tumours used in this part of the study are the results of US scans conducted on women recruited into the IOTA study (Timmerman, Van Calster et al. 2010). All women in this study underwent surgical removal of the tumours between November 2005 and November 2013 with a known histological diagnosis. They all signed a written informed consent form to allow the use and analyse the data for research purposes (Timmerman, Van Calster et al. 2010). This study was granted ethical approval by the University of Buckingham's School of Science & Medicine. We retrieved a total of 187 ultrasound images of 177 anonymous patients from the IOTA database (Astraia software gmbh, Germany) at the Department of Gynaecological Ultrasonography, Campus Gasthuisberg, KU Leuven, Belgium. This means that for ten of the patients we obtained additional 10 images with another representative region of interest (ROI) that reflected the final histopathology.

Each image is a 2D B-mode ultrasound scan image of the surgically removed tumours. The used image for each tumour, selected by Dr Jeoren Kaijser of the KU Leuven Hospital on subjective impression, was the most representative of the final histopathology. Out of the 187 US images, 112 show benign tumours and 75 malignant tumours. Histology results for all tumours corresponding to the images available in Table 3.1.

Table 3. 1: Histopathology of ovarian tumours included in our training and test groups.

	Histopathology	N (total N=187)
Benign(n=112)	Mature teratoma	23
	Endometrioma/endometriosis	15
	Mucinous cystadenoma	23
	Functional cyst	5
	Ovarian fibroma	6
	Serous cystadenoma	21
	Serous cystadenofibroma	13
	Other Benign	6(1 tubal abscess, 1 Brenner tumour, 1Multilocular peritoneal inclusion cyst MPIC, 1 Mucinous cystadenofibroma, 1 subserous adenomyoma, 1 hydrosalpinx)
Malignant (n=75)	Borderline mucinous tumour	15
	Borderline serous tumour	6
	Serous cyst/adenocarcinoma	28
	Mucinous cyst/adenocarcinoma	3
	Endometrioid adenocarcinoma	6
	Other ovarian cancer	17 (1 Lymphoma, 7 metastatic tumours(3 intestinal, 1 breast, 1 pancreatic, 1 gastric, and 1 lung cancers), 1 leiomyosarcoma, , 1 stromal tumour, 1germ cell tumour, 2 clear cell carcinomas, 2 carcinosarcomas, 1 immature teratomas, 1 endometrial cancer)

The data set reflects a wide range of variations in tumour types and shapes/components from a large number of patients. This is to ensure the developed automatic diagnosis solutions can deal with these various forms of tumours.

3.4.2 Static B-mode Ultrasound Images of Gestational Sac

The ultrasound images used in this component of our research project were obtained in two batches. The first dataset batch contains 94 ultrasound images: 79 images are PUV cases and

15 images of known miscarriage cases. The second independently sampled dataset contains 90 images among which 78 images are PUV cases and 12 are of known miscarriage cases. In total, the experimental dataset consists of 184 images of which 157 are images of PUV cases and 27 images are of miscarriage cases.

All images were taken at various points of time in the first trimester of pregnancy, collected and labelled by Dr Jessica Farren in the Early Pregnancy Units, Imperial College Healthcare Trust, London, UK. These images were collected between 17/6/2013 to 16/6/2014. The two batches of images are taken from different machines and at the different time. Each image consists of two views of a GS from both sagittal and transverse planes. All images are provided with the MSD values based on manual measurements and the diagnosis results. All patients signed a written informed consent form to allow the use and analyse the data for research purposes. This study was granted ethical approval by the University of Buckingham's School of Science & Medicine.

3.5 Classifier and Classification Methods

Classification is a main task in machine learning. It is a process of discrimination between two or more different classes by labelling each similar set of data with a certain label to differentiate it from other classes. Image classification is performed in two stages, *training and testing.* a set of training sample images with known class labels are used to train a classifier such as k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), decision tree (DT), etc. More details about various types of classifiers can be found in (Fix and Hodges Jr 1951). Once the classifier is trained, it is used to predict the class of an image. The performance of the classifier can be judged by classifying a test image of known class: if the predict class is the same as the known class of the test image, the classification is accurate; otherwise it is not.

Our classification framework consists of selecting an appropriate classifier(s) that fit our requirements, adopting appropriate evaluation protocols, and investigate fusion strategies. In addition, our exploration of effective classification methods is motivated by two reasons. First, we want to find classifiers that are effective to be applied to the extracted features. Second, we want to apply classifiers that are effective for each of the two problem areas, i.e. the diagnosis of ovarian tumour and identification of miscarriage cases in early pregnancy. The feature spaces for the two problem areas are very different. The dimensionality of the feature space for ovarian tumour diagnosis is high, while the dimensionality of the feature

space for miscarriage identification is very low (See Chapter 5 for more discussions on the proposed features). After an extensive review of the existing classification methods, we focus on two specific solutions: k-Nearest Neighbour (kNN) and Support Vector Machines (SVM). Both classification methods are directly applicable to the types of features we extracted from ultrasound images, i.e. vectors of numeric quantities, without further pre-processing to the feature vectors. Besides, according to the research, kNN classifier with the Euclidean distance should be suitable for a low dimensional feature space (Beyer, Goldstein et al. 1999). An SVM classifier, on the other hand, should be suitable for a feature space of high dimensionality (Han and Kamber 2001). More details about these two classifiers will be explained next.

3.5.1 k- Nearest Neighbour (kNN) Classifier

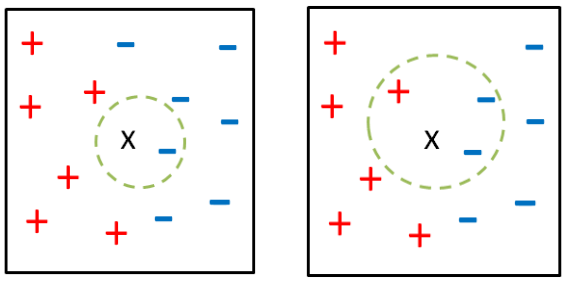
The kNN method was first described in the early 1950s (Fix and Hodges Jr 1951). It has since been widely used in the area of pattern recognition. It is a simple algorithm that stores the examples from the training set as templates and classifies a new unknown test case based on a similarity measure (e.g. a distance function) between the new case and every template. The test case is classified by assigning the class label of the majority among the k nearest neighbouring templates of the new case. "Closeness" can be defined in terms of a distance metric, such as Euclidean distance (Webb 2003). The kNN algorithm for classification is outlined as follows:

1. Compute the distance between test case $X = (x_1, x_2, \ldots, x_n)$ and each training template $Y = (y_1, y_2, \ldots, y_n)$ using a distance measure such as the Euclidean distance D_2:

$$D_2(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.2}$$

2. Sort all distances in ascending order and identify *k* nearest neighbors of X: $Y_1, Y_2, \ldots, Y_k$, where $k \geq 1$
3. Assign X the class label of the majority among $Y_1, Y_2 \ldots Y_k$,

Figure 3.5 illustrates an example of the kNN algorithm in action. A test case X has its class unknown. When k = 1 as shown in Figure 3.4 (a), the nearest neighbor of X is of the negative class, and therefore, the test case is assigned to negative. When k = 3 as shown in Figure 3.4 (b), the majority voting by three nearest neighbours also determines that the test case is of the negative class by two negative votes against one positive vote (Han and Kamber 2001).



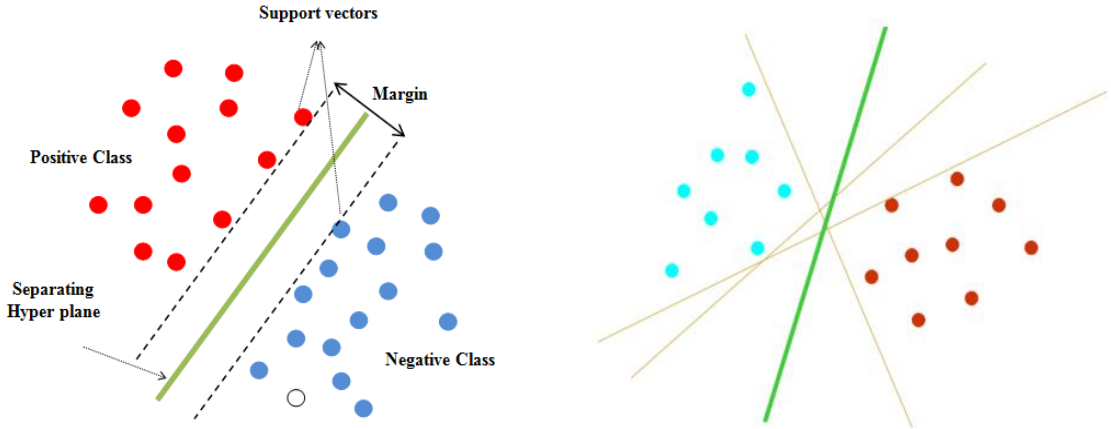
(a) 1- nearest neighbour (b) 3- nearest neighbours

Figure 3. 4: Illustration of the working of kNN classifier

3.5.2 Support Vector Machine (SVM)

The original SVM algorithm was invented by Vladimir N. Vapnik and was proposed by Vapnik and Corinna Cortes in (Cortes and Vapnik 1995). SVM is based on the concept of decision planes that define decision boundaries. A decision plane separates a set of objects of the positive class from those of the negative class. An illustrative example is shown in Figure 3.5 (a). The separating line defines a boundary between the objects of the two classes (BLUE and RED). Any new object falling to the right side of the line is classified as BLUE, or to the left side as RED. The white circle object in the example is therefore classified as BLUE. The example in the figure is a typical example of a linear SVM classifier, i.e. the separation boundary between objects of their respective groups is a straight line. This separation line between two classes in higher dimensions is known as a hyperplane.

There are many possible hyperplanes that can separate the data of two classes, but there is only one that maximises the margin (maximises the distance between the plane and the nearest data point of each class). Such a hyperplane is known as the optimal separating hyperplane (Han and Kamber 2001). Figure 3.5 (b) shows different lines with the best separating hyperplane marked in green. It is optimal because it maximises the differences between the two classes. Only the features on the margin boundaries are used to discriminate between the two classes; these features are called “support vectors”.



(a) (b)

Figure 3. 5: (a) Schematic representation of the principle of SVM. SVM tries to maximise the margin from the hyperplane in order to best separate the two classes (red positives from blue negatives) (b) Optimal separating hyperplane

The algorithm to classify a new case goes through the following steps:

1. Dataset was split into training and testing subsets, and then SVM was trained on the training data set regarding the labels of the classes, to find the best hyperplane based on margin maximisation between the two classes:

   SVM train (Training set, Labels of training set)

2. Find the support vectors after standardize all training points according to the mean (μ) and the standard deviation (σ) of all training samples for each feature component domain:

$$Si = \frac{xi-\mu}{\sigma} \tag{3.3}$$

3. Calculate the score (Sc) of the test case x based on Lagrangian formulation [13][12]

$$Sc(x) = \sum_{i=1}^{n} \alpha_i \, k\,(Si, x) + b \tag{3.4}$$

where Si is a support vector, α_i is the weight of Si, n is the number of support vectors, k is a kernel function. In the case of linear classifier, the kernel function k is simply the inner product $<s_i, x>$ and the bias b.

4. If $Sc \geq 0$, then x is classified as positive, otherwise it is classified as negative.

Both kNN and SVM are popular classifiers and have been successfully used in many fields of application, such as bioinformatics, text, image classification, face recognition and medical image recognition (Cristianini and Shawe-Taylor 2000).

3.6 Fusion Methodology

A common practice in pattern recognition, and particularly when different recognition schemes are known to solve the problem but none achieves acceptable/desirable level of accuracy, is to fuse/combine one or more of these schemes with the hope that they can complement each other and collectively improve accuracy of recognition. Over the years, the benefits of fusion have been observed in a variety of pattern recognition applications from different disciplines and as result fusion has become one of the major research topics in classification and pattern recognition. Accordingly, fusion is an essential component of our adopted research framework.

As we explained in Chapter 2, the fusion of diagnosis decisions is a common practice in clinics, particularly for ovarian tumour diagnoses. Moreover, the levels of confidence associated with the diagnosis decisions are reflected by assessing a risk factor of malignancy as explained in details in Chapter 2. Fusion often leads to improving diagnosis confidence and better and more accurate eventual diagnosis decision. Occasionally, fusion may also create conflicts between the classifiers that lead to a rather not-so-certain situation where a further investigation is required. Therefore, the scheme of fusion in this thesis is largely motivated by and hence close to this clinical practice, details about our novel work can found in Chapter 6, Section 6.5.

Various approaches for fusion have been successfully implemented for different applications over the past years, such as recognitions of fingerprints, irises, facial images, and hand geometry (Hangiandreou 2003, Bushberg and Boone 2011). In general, four levels of fusion, i.e. pixel level fusion, feature level fusion, scores level fusion and decision level fusion, have been investigated (Bushberg and Boone 2011), and summarised as follows:

a. **Pixel level fusion:** At this level, pixels of individual source images are fused, usually through some forms of processing to the source images (e.g. edges extraction or texture analysis) (Dhawan 2011), into an abstract meta-image before features useful for image classification are extracted from the meta-image.
b. **Feature level fusion:** This level of fusion combines different feature vectors obtained by multiple feature extraction algorithms. Often, the feature vectors are concatenated into a

single feature vector or some forms of feature aggregation are applied (Bushberg and Boone 2011, Dhawan 2011).

c. **Scores level fusion:** This fusion scheme combines classification scores into a final score that further determines the class. In general, score level fusion can be divided into three categories: (a) Transformation-based score level fusion, (b) Classifier-based score level fusion, and (c) Density-based score level fusion (Michailovich and Tannenbaum 2006).

d. **Decision level fusion:** This fusion approach combines classification decision outcomes, which can be made by different classifiers or by using different feature vectors, into a final decision (Bushberg and Boone 2011).

3.7 Experimental Protocol and Performance Evaluation

3.7.1 Evaluation Protocols

Classification is achieved in two stages: training (input samples with their extracted features and known class labels are used to learn in the model from the data) and testing (determining classes of test examples with class labels and calculating the accuracy of the predicted classes by the model). Given a set of data, a protocol/policy is needed to decide how to divide the data into training and testing sets. Among such protocols, one can select the:

- Hold-out. This method randomly divides the data into two subsets A and B. The classifier is trained with the subset A and tested with the subset B first, and then trained with B and tested with A (Webb 2003). The accuracy is taken as the average of the accuracies of the two tests.
- *k*-fold Cross Validation. This method divides a given data set into *k* sub-sets with the equal number of examples. A classifier is tested on each subset and trained on the rest of the data. This training and testing process is repeated *k* times with different testing sets, and the average performance of the N tests is the classifier performance (Duda, Hart et al. 2001, Webb 2003).
- Leave-One-Out method. Is a special case of the N-fold cross validation. The data set is split into a training set of size N-1 and a testing set of size 1, and the classifier is trained and tested. This process is repeated until all the data in the data set are tested (Webb 2003).

In this study, we used a stratified k-fold cross-validation, which means applying the leave-one-out strategy to utilise the use of training examples. More details about this protocol are below.

3.7.1.1 Evaluation Protocol for Ovarian Tumour Diagnosis Schemes

In Section 3.4, we described in detail the experimental data set of ultrasound image of ovarian tumour that has been used in this study. To address the class imbalance problem between Benign (112 cases) and Malignant (75 cases) in the dataset and to develop a fair classification model for both cases, we randomly down-sampled 50 images of Benign and 50 images of Malignant tumours, totalling 100 images for training and testing. The sampling was performed without replacement using the Randsample function in Matlab (50 Benign with Randsample (112, 50), and 50 Malignant Randsample (75, 50). For evaluation of performance, we employed a stratified leave-one-out cross validation, i.e. in an iterative process, one partition consisting of two samples (one Benign and one Malignant) was taken as the test examples and the rest for training the SVM (Figure 3.6). We repeated this process 15 times with a different random selection of 100 images (50 Benign and 50 Malignant) to reduce the random effect and make the performance measurement more reliable. Then at the end, we take the averages of accuracy, sensitivity and specificity respectively for the 15 rounds.

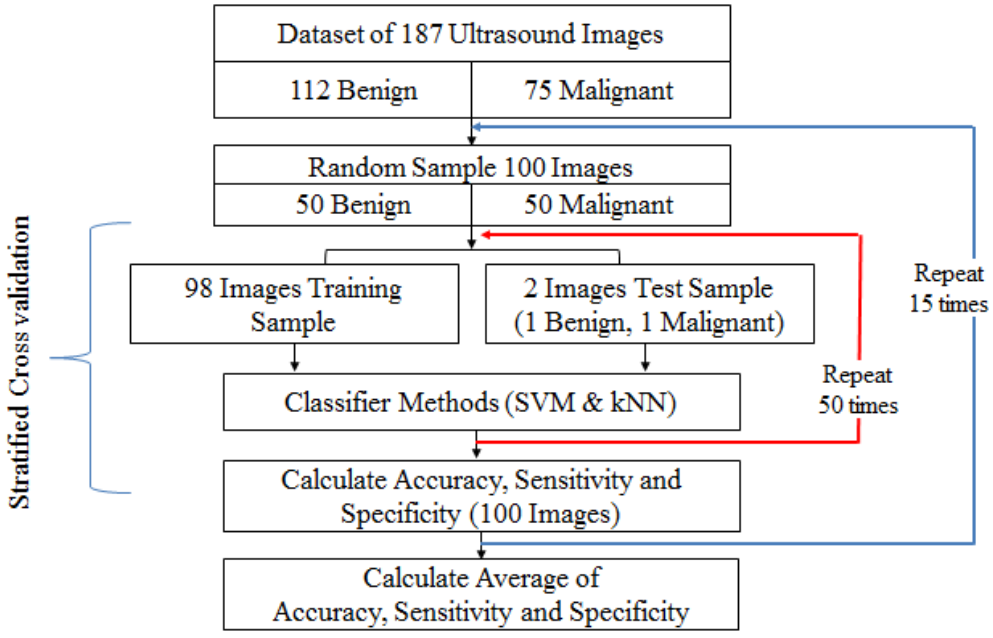


Figure 3. 6: Randomised balanced cross validation for selecting training and test groups – Flow chart

3.7.1.2 Evaluation Protocols for Miscarriage Identification Scheme

As mentioned earlier the ultrasound miscarriage related images used in this thesis were obtained in two batches. In the medical science field, testing a new model of diagnosis normally follows a split testing strategy (as to cross-validation for machine learning). The model is first developed using a medical centre's own data and then tested with an independently sampled set of internal test data from the same medical centre. This is then followed by using an external set of independent test data collected from a different medical centre, We therefore, tried to follow both approaches of testing from the medical science field (without the external testing part because of the constraint on data source) as well as from machine learning, and hence conducted two different sets of experiments to test the performance of the developed miscarriage identification scheme. Unfortunately, we could not do so for the ovarian tumour diagnosis problem because the sample images were from the IOTA database where the images are from multiple medical centres).

So in the first set of experiments, we used the first set of 94 images (79 images are PUV cases and 15 images of miscarriage cases) as a training set and the newest set with 90 images (78 images are PUV cases and 12 are of miscarriage cases) as an internal test set. To overcome the class imbalance problem between the PUV and miscarriage cases in the training set, we employed a sampling strategy in rounds, similar to that as explained in Section 3.7.1.1. In each round, we randomly selected 15 images from 79 images of PUV cases and combined them with the 15 images of miscarriage cases to form a training sample of 30 images. We then train a classifier using the 30 images, but we test the performance of the classifier using the whole of the internal test set. We again repeated the training with different random samples of 30 images and testing with the internal test set 15 rounds and then collected the averages of accuracy, sensitivity and specificity respectively. Figure 3.7 illustrates this testing protocol.

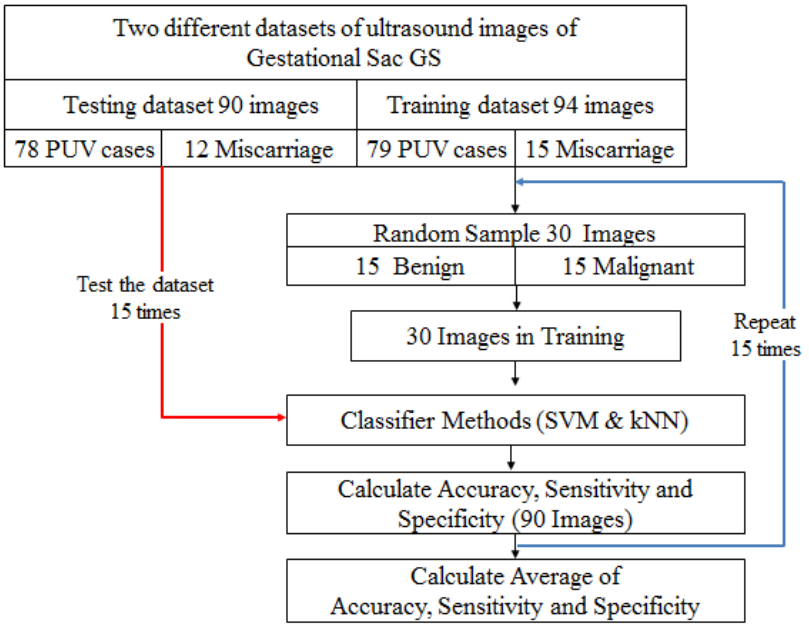


Figure 3. 7: Randomised balanced cross validation process of selecting the training set – Flow chart

In the second experiments, we adopted a typical machine learning test option by combining all datasets in a single dataset of 184 images (157 PUV cases and 27 miscarriage cases). A stratified cross-validation was employed as the validation protocol for this experiment. We also used a similar sampling strategy to resolve the class imbalance problem. Namely, a random sample was drawn 15 times. For each sample, 25 randomly selected (out of 157) PUV images and 25 randomly selected (out of 27) miscarriage images were chosen to feed into the cross-validation. We then determine the average of accuracy, sensitivity and specificity. Figure 3.8 clearly explain the testing protocol for this experiment.

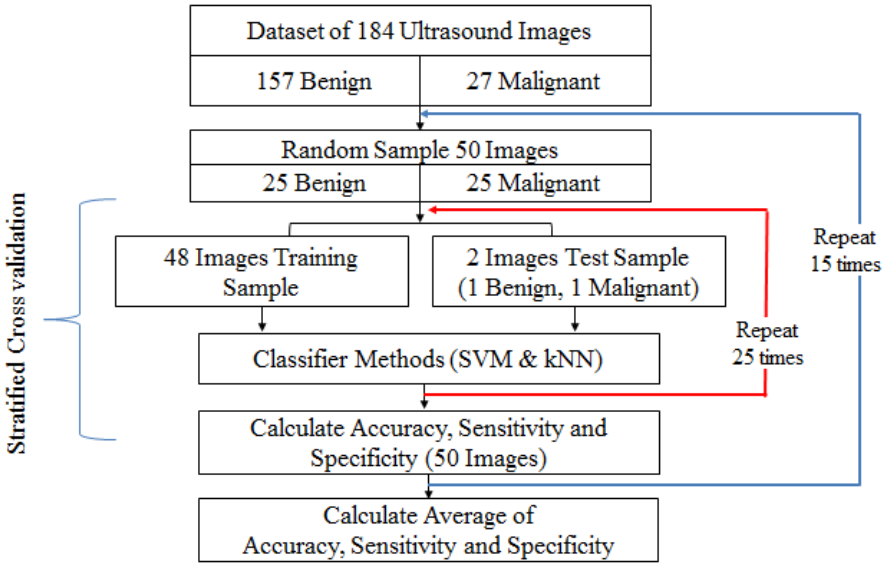


Figure 3. 8: Randomised balanced cross validation for selecting training and test groups – Flow chart

3.7.2 Performance Evaluation

Evaluating the performance of any procedure proposed to solve a pattern recognition scheme or a classifier would require the adoption of standard evaluation measures. We first use the overall classification accuracy. The accuracy rate refers to the proportion of the number of successfully classified test cases to the total number of tested cases. The formula for calculating the accuracy rate is presented as follows and the terms used are explained in Table 3.2:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{3.5}$$

Table 3. 2: Description of Performance parameters obtained in the binary classification test.
True positive **TP**: Sick people correctly diagnosed *(correctly identified)*
False positive **FP**: Healthy people incorrectly identified *(incorrectly identified)*
True negative **TN**: Healthy people correctly identified *(correctly rejected)*
False negative **FN**: Sick people incorrectly identified *(incorrectly rejected)*

We also calculate two measures of the performance of a binary classification test: Sensitivity and Specificity, which are commonly used to evaluate clinical tests.

- **Sensitivity**: It refers to the rate of correctly identifying those patients who actually have the disease.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3.6}$$

- **Specificity**: It refers to the rate of correctly identifying those patients who actually do not have the disease.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \tag{3.7}$$

As mentioned above, our frameworks for measuring the performance of the proposed schemes include the introduction of an innovative measure of the strength/reliability of each single diagnostic decision made by the adopted classifier. See Chapter 6 for more details.

3.8 Summary

In this Chapter, we presented an overview of our research framework for automatic analysis and diagnosis of gynaecological abnormalities from Ultrasound scan image. The framework outlined the research to be detailed in the following chapters about each of the steps in the process starting with enhancement, followed by feature extraction and classification. We also included the framework for the experimental test validation procedure. In particular, we described the datasets used in the following chapters of the thesis for both medical problems, i.e. scan images of different types of ovarian tumour and two datasets of gestational sac. Lastly, the experimental protocol and performance evaluation used in the following chapters were explained in details. In the next chapter, we will present our research work on automatic pre-processing enhancement methods.

CHAPTER 4

ENHANCING AND SEGMENTING GYNOLOGICAL ULTRASOUND IMAGES

Ultrasound scanning is the most commonly used imaging system in the gynaecological area. The main purpose behind the scanning is to detect a region of interest ROI such as a tumour to monitor a specific type of disease or diagnose cases of miscarriage from GS in an early stage of pregnancy. Having introduced the research problem dealt with in this thesis and discussed the background material relevant to our project, we now embark on the first component of our investigations. This chapter focuses on pre-processing and segmentation tasks relevant to computational aspects of gynaecological ultrasound images in preparation for the subsequent analysis steps. In particular, the main contributions of this chapter are:

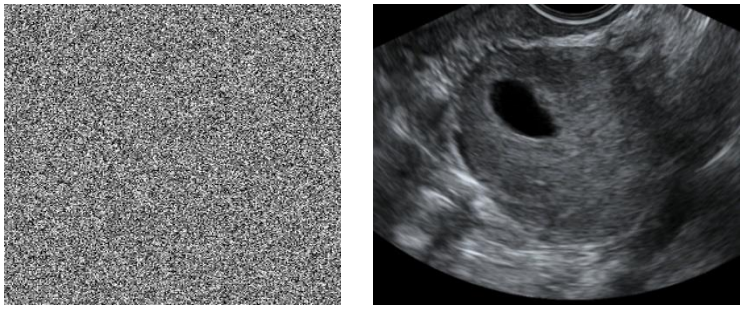
- Proposing a new method for enhancing ultrasound images of ovarian tumours based on non-local mean filter followed by a negative transformation and the absolute difference function.
- Automatic enhancing of ultrasound image of the gestational sac based on a mean related operation.
- Automatic segmentation of ultrasound image of the gestational sac based on Otsu thresholding followed by a set of image processing procedures to remove all non-sac binary objects

In section 4.1, we briefly explain a specific type of noise that affects the ultrasound images namely, the speckle noise. We also highlight how to exploit different image processing techniques to eliminate the effect of this type of noise in order to enhance the quality of the anatomical information content of ultrasound images. The section concludes by explaining the segmentation task and review existing work in the literature on extracting the regions of interest in medical ultrasound images. Enhancement/segmentation objectives vary according to the requirements associated with the two types of ultrasound images under investigations. Section 4.2 describes our work on enhancing and segmenting the gestational sac ultrasound images, while section 4.3 is devoted to our work for enhancing ovarian tumours in ultrasound images. Section 4.4 summarises the work carried out in this Chapter.

4.1 Literature Review

In general, there are many artefacts (including noise) that affect the quality of medical imaging modalities with possibly adverse impact on the automatic processing/analysis of these images for the intended purposes. Each type of modalities is often corrupted by a different type of noise, which tends to reduce diagnosis accuracies. For instant, Rician noise is a specific noise that is known to corrupt MRI imaging system. The noise belongs to the additive type of noises which makes the quantitative measurement difficult. Salt & Pepper noise affects CT scan images. This type is impulse type of noises. For more details, see (Dangeti 2003) and (Gonzalez, Woods et al. 2004).

Despite the wide use of ultrasonography imaging in medicine, the interpretation of such images is considerably dependent on visual examination by domain experts, and hence high image quality is an essential requirement. Ultrasound images are known to be corrupted by a special type of noise, known as "Speckle noise" consisting of artefacts caused by interference of energy from randomly distributed scattering (Zhu, Ni et al. 2009). The noise belongs to the multiplicative noise type. Figure 4.1 shows the speckle noise and its effect on ultrasound images of ovarian tumour. Speckle noise has a negative impact on image quality by hiding and blurring important details particularly on edges and, therefore, affecting the later image segmentation and other post-processing operations, and eventually may reduce the diagnostic value of the image (Loizou and Pattichis 2008, Zhu, Ni et al. 2009). Speckle noise reduction is a particularly important requirement for automatic processing and analysis of ultrasound images.



(a) (b)

Figure 4. 1: Illustrates the speckle noise and its effect on ultrasound images of ovarian tumour (a) Example of Speckle noise (Dangeti 2003) (b) Ultrasound image of ovarian l tumour corrupted by a speckle noise

In automatic biomedical image processing and analysis, it is very important to eliminate or at least reduce the effect of noise to minimise its negative impact on the later stages i.e. the

segmentation and features extraction stages, and to increase diagnosis (classification) accuracy manually or automatically. There are two main factors that tend to affect the ultrasound scanning process and image capturing. First, there is a wide range of ultrasound machines with possible different parameter settings. Second, sonographer's expertise and experience in capturing the US images can vary significantly. Both factors may result in US images of different qualities and inclusion of noise levels. There are different de-nosing and enhancing techniques that have been used to improve medical imaging modalities including ultrasound images with varying degrees of success (Rose and Allwin 2013). Unfortunately, there is no specific single method for removing the noise and enhancing images that works on all types of ultrasound scan images. To eliminate the effect of the noise, computer scientists should be careful in selecting and choosing their de-nosing techniques as some may affect the anatomical information inside ROI. Therefore, an experienced and expert radiologist (sonographer) needs to be consulted on the best de-noising approach by evaluating the resulting US images. In the following sub-sections, we shall explore existing methods and techniques in de-noising and enhancing US images.

4.1.1 Existing Methods for Ultrasound Image De-noising and Enhancing

There are several published works in de-noising/enhancing ultrasound images for different types of medical diseases, where various techniques are used to enhance the image for automatic segmentation or facilitate visualising image details for manual diagnosis (Malathi and Shanthi 2010, Thaipanich and Kuo 2010). Image de-noising is a known image processing problem that is usually resolved by applying noise filters. Many de-nosing filters have been developed in both spatial and frequency domains (Buades, Coll et al. 2004). Different speckle noise reduction filters have also been proposed in the literature. The commonly used linear low-pass filters, such as mean filters, are not suitable for reducing the speckle noise of US images since they also remove the high frequencies which result in smoothing out image edges (Xie, Jiang et al. 2006). The median filter and its adaptive version can effectively suppress speckle but fails to preserve other useful details (Xie, Jiang et al. 2006). Czerwinksi et al. proposed the stick filter (Czerwinski, Jones et al. 1999) for reduce the speckle as well as improving the edge information. This filter has been used to de-noise prostate ultrasound images. Despite the limited success, however, a recent survey conducted by (Zhu, Ni et al. 2009) summarised and highlighted some major disadvantages of the existing works. Firstly, each filter is sensitive to a specific model of noise. Secondly, most filters succeed in reducing noise in smooth and background areas, but not sufficiently well with enhancing image

features such as edges. Thirdly, most existing filters are not adaptive and have the "smoothing effect" on edges and other distinct image features. Therefore, noise removal for medical ultrasound images is still a challenging task.

Over recent years, a more effective approach of applying a noise removal filter followed by special image enhancement operations started to emerge. In (Pathak, Haynor et al. 2000), speckle noise reduction was achieved for prostate ultrasound images by using stick filter followed by anisotropic diffusion filter to smooth the images for successfully segmenting prostate tumours, leading to improved results in diagnosis. Further improvement of automatic segmentation of prostate US images using stick filter and top-hat transform was reported in (Rafiee, Salimi et al. 2008) where the result achieved was very close to the manual segmentation. A simple method using histogram equalisation was also applied to enhance the placenta ultrasound images (Malathi and Shanthi 2010), which was effective enough for improving the quality of the image by highlighting the ROI for accurate segmentation. A new de-nosing approach using the so called Non-local means filter was proposed in (Buades, Coll et al. 2005). This filter replaces each pixel by the weighted averaging of all pixels in the image that have similar neighbourhoods. This method aims to remove the noise while keeping all meaningful information. It aims to take advantage of the redundancy and self-similarity of the image. Adaptive Non-local means filter was proposed in (Thaipanich and Kuo 2010) to effectively improve ultrasound images. The method also was successfully applied to MRI images of removing the Rician noise (Zhang, Hou et al. 2014, Yang, Fan et al. 2015).

There is also a great deal of research in frequency domain for de-nosing ultrasound image of follicles. The Gaussian low-pass filter was used on follicle in ultrasound images followed by histogram equalisation (Hiremath and Tegnoor 2010). The main aim of that work was to highlight the follicles for detection at the later stage. The method was further improved in (Hiremath and Tegnoor 2013) by using a negative transformation on the enhanced image, which improved the detection significantly.

Research work in the area of GS ultrasound image de-nosing and enhancing is very limited. In (Chakkarwar, Joshi et al. 2010), contrast enhancement was used followed by a combination of smoothing low pass filter and wiener filter. The result was helpful in segmenting the GS. From the literature on ultrasound image de-noising and enhancing, we noticed that different filters and methods were used for different purposes, i.e. just removing

the noise from the image or de-noising for the purpose of highlighting the ROI to be prepared for segmentation stage or de-noising for the reason of highlighting the anatomical structure inside the ROI to be prepared for feature extraction step. There is no a specific method that works for all types of ultrasound images because each image comes from a different source machine with the different setup and captured by different sonographer who tends to have the different level of experience.

4.1.2 Ultrasound Image Segmentation

In the context of image analysis, segmentation is the process of separating the distinguished or interesting image objects from the image background, i.e. partitioning the image into meaningful regions and then selecting the wanted region known as Region of interest (ROI) (Setarehdan and Singh 2012). Image segmentation is an important step that can make the image analysis easier. In medical imaging applications, ROIs are often extracted either manually by experts or automatically using image segmentation methods before analysing the image for the purposes that were captured for. Figure 4.2 show examples of B-mode ultrasound images of ovarian tumours, and manually segmented ROIs, i.e. tumours highlighted from the background by yellow border lines.

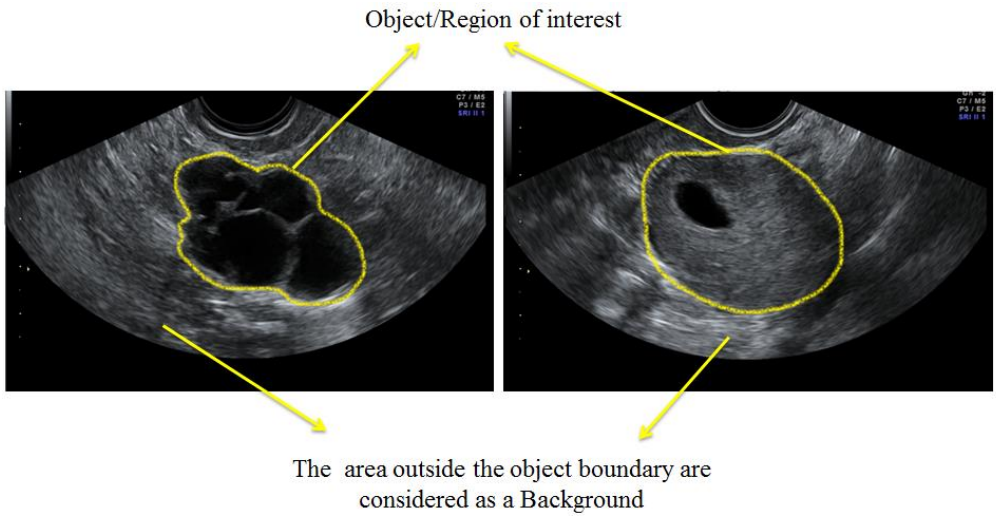


Figure 4. 2: Ultrasound images of ovarian tumours and highlighting the ROI

Segmenting an ultrasound image is strongly influenced by the quality of the image (Özgen 2011). One of the main difficulties is that in some cases the ROI and the background have the similar texture which makes automatically finding the ROI very difficult (Zhang 2006). Various traditional methods and innovative approaches have been developed. Straightforward binary segmentation methods such as “thresholding” and “region growing” may be preferred

due to the fact that they are both easy to implement and fast (Manay and Yezzi 2003, Özgen 2011). However, such methods normally have rather limited success in dealing with complex US images of tumours.

Several approaches were proposed to segment the ROI in ultrasound images, for detecting a different kind of diseases. For example automatic detection of the liver tumour in ultrasound images was proposed using Otsu thresholding after images were enhanced (Hiransakolwong, Hua et al. 2003). In (Al-Bayati and El-Zaart 2013), Otsu thresholding was also used to segment the breast tissue in ultrasound images and good results were reported in detecting normal breast tissues.

A novel method for automatic prostate segmentation in ultrasound images was presented in (Rafiee, Salimi et al. 2008), for extracting the boundary of the prostate. This method works by selecting seed point inside the prostate for an active contour procedure. A number of experiments were conducted to validate this method and the results showed that the algorithm output was very close to the manual method of extracting the prostate boundary by experts.

A number of techniques have been investigated in the area of follicle detection in ultrasound images of ovaries. Follicles appear as dark regions in ultrasound images. Therefore, the goal of the segmentation process is to identify most of the dark regions in an ultrasound image. In (Harrington 2007), a method for follicle detection was proposed that works in two steps: 1) segmenting of all dark regions in the image and 2) removing the non-follicle regions from the set of segmented regions. The segmentation step uses geometric active contour models (snakes). However, the performance of that algorithm in detecting small follicles (diameter smaller than 3 mm) could be adversely affected by the poor quality of ultrasound images. Determining suitable parameters for automatically initiating the construction of snakes is a challenge that remains unresolved. Recently a novel method for follicle detection by employing active contour method was published in (Hiremath and Tegnoor 2010), because ovarian follicles are oval shaped compact structures, which resemble the ellipse, characterized by the seven geometric features, namely, the area, the ratio of major axis length to minor axis length, the compactness, the circularity, the tortuosity, the extent and the centroid. Experimental results demonstrate the effectiveness of this method. Another approach to follicle segmentation in ultrasound image was proposed in (Hiremath and Tegnoor 2010) that apply canny edge detection followed by morphological dilation operator to fill broken edges. The results of experimental tests demonstrated the efficiency of follicle

segmentation by this algorithm when compared with the inferences drawn by the medical expert.

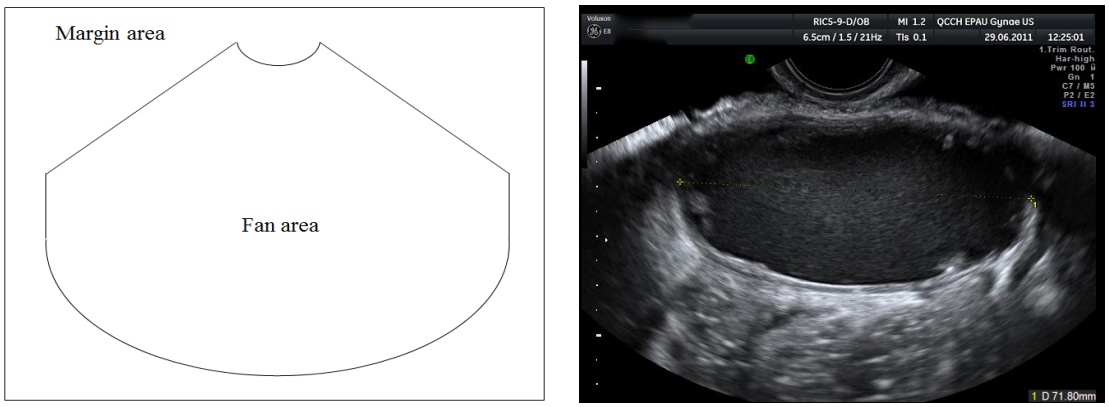
Unfortunately, research in the area of automatic segmentation of gestational sacs in ultrasound image is somewhat limited. In (Chakkarwar, Joshi et al. 2010), Otsu thresholding was applied on a rather small dataset of 12 ultrasound images, and the gestational sacs were segmented successfully after enhancing the images. In (Zhang, Chen et al. 2011), an algorithm for detecting the GS from a video was proposed, using the AdaBoost method to detect the GS from each frame. The algorithm was tested on 31 videos and achieved a GS detection rate of 87.5%.

To the best of our knowledge, there are no other specific techniques reported in the literature to segment ultrasound images of ovarian tumour. In fact, it can be argued that ovarian tumour segmentation in ultrasound images is still an open problem and a challenging task due to some difficulties related to the tumour structures (Sohail, Rahman et al. 2010, Hamid 2011).

4.2 Enhancement and Segmentation of US Images of Gestation Sac (GS)

In this section, we describe in details our proposal for enhancing and segmenting pregnancy-related US images where the region of interest in this case is the GS. Although such US images are also corrupted by noise, the fact that the GS is a single object of interest whose internal intensity distribution can easily be distinguished from its surroundings removes the need for de-noising the image first. Thus, GS segmentation requires somewhat simple enhancement procedures.

In general, ultrasound image contains two parts: the fan shaped area, which represents the image of interest obtained by an ultrasound scan, and the marginal area, see Figure 4.3 (a). The marginal area contains information about the patient (blocked for anonymity), date and time that the image was taken and the ultrasound machine setting. Figure 4.3 (b) shows ultrasound image with both parts. The Fan area is important part which contains ROI.

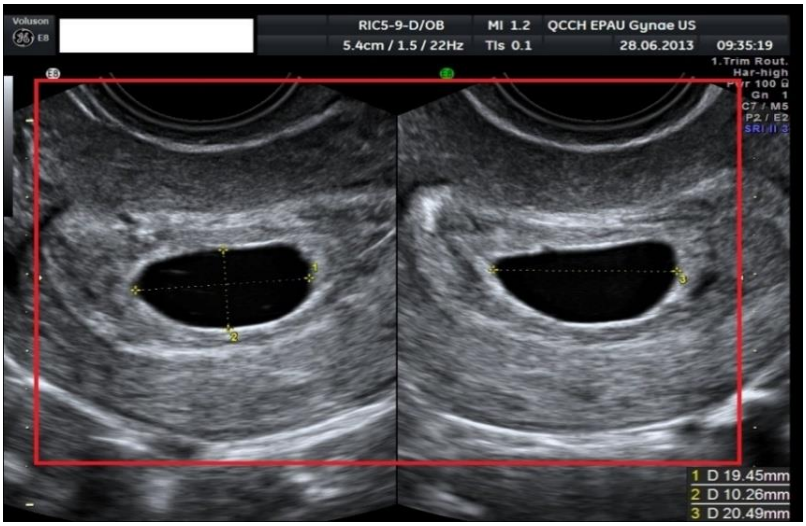


(a) (b)

Figure 4. 3: Example of ultrasound image with fan and margin areas (a) The boundaries of the fan and margin areas (b) Ultrasound image present both areas

4.2.1 Image Preparation

Figure 4.4 presents an example of the ultrasound image of the GS taken in respective sagittal and transverse planes. The red rectangle represents the main fan area. The GS is the dark region in the centre of the fan area. The area outside the red rectangle is the margin area. In the Fan area, three manual measurements of the GS size (in millimetres) are marked by yellow dotted lines, and their estimated values are displayed in the margin area.



Sagittal plane view Transverse plane view

Figure 4. 4: Ultrasound images of GS in Sagittal and Transverse planes

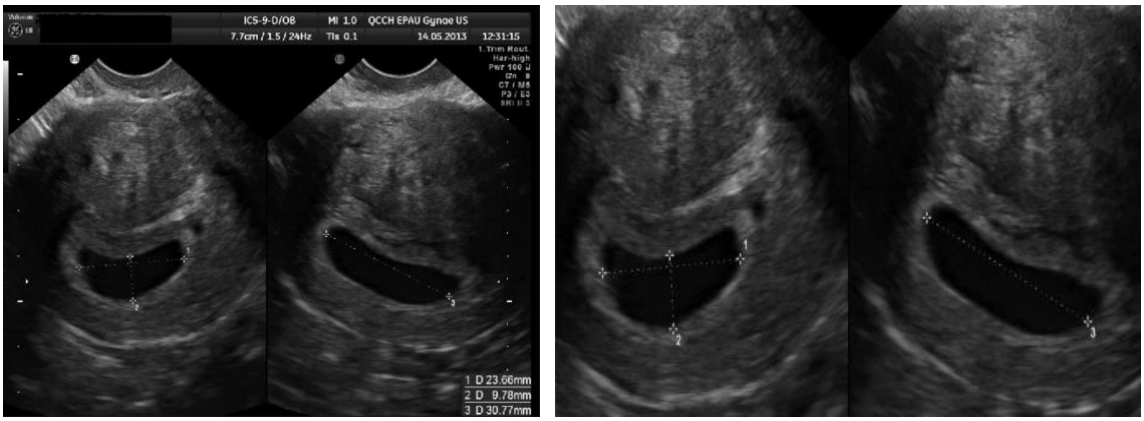
Before the process of GS segmentation starts, the margin areas of both views need to be removed. We used the imcrop function in Matlab with a fixed position vector parameter (30, 150, 900, and 500). The cropped image is as shown in Figure 4.5 (a). Then we separate the two views from the middle of the image. The resulting two images are shown in Figure 4.5(b).

4.2.2 Image Enhancement

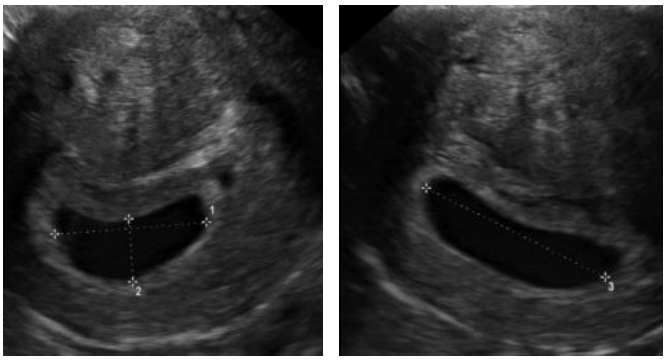
Ultrasound images of the GS are typically dark, causing difficulties in segmenting the GS. However, the GS is somewhat darker than its surrounding. There are various techniques to enhance such images. We use a simple method to make the bright areas brighter while the GS become more highlighted, using a heuristically defined simple filter:

$$J(i,j) = \begin{cases} 2 \times I(i,j) & if\ \mu \geq 55 \\ 4 \times I(i,j) & otherwise \end{cases} \quad (4.1)$$

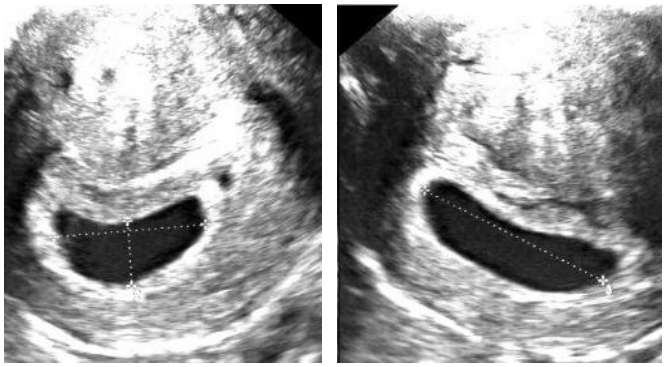
Where *I(i,j)* is the intensity value of the pixel at (i, j) position, and μ is the mean of all pixel intensity values. Unlike histogram equalisation, this simple pixel value transformation gives more weight to the less dark pixels in the surrounding areas of the GS by stretching them over the whole greyscale range. The main outcome of this enhancement is that it highlights the GS as a single dense region which is easy to segment, see the highlighted area of the GS in Figure 4.5 (c).



Ultrasound image of GS (Both plane views) (a)



(b)



(c)

Figure 4. 5: Shows steps of image subtraction, separation and enhanced (a) The Cropped image (b) Separated both planes (c) The Enhanced images in both planes

4.2.3 GS Segmentation

The GS segmentation stage involves a series of operational steps that overcome the effects of the blurry and noisy nature of the surrounding area near the sac as depicted in the block diagram shown in Figure 4.6.

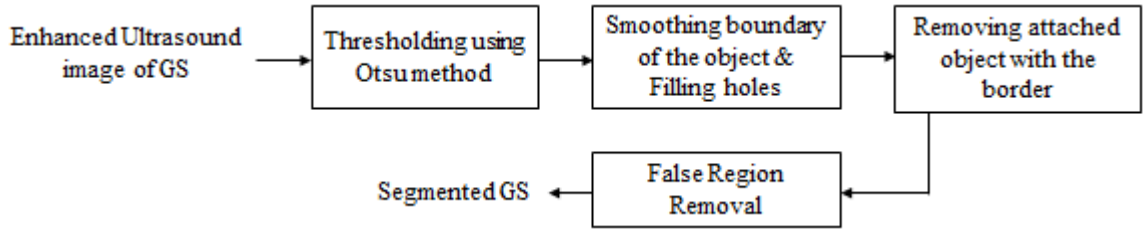


Figure 4. 6: The block diagram of the process of segmenting GS

Step 1: Otsu Thresholding (Otsu 1975). This thresholding technique is widely used in medical image and specifically in ultrasound image segmentation because of effectiveness as well as simplicity of implementation. The idea of the Otsu threshold is that the object of interest and background in a given image *f(x,y)* have intensity levels grouped into two parts. Basically this method search for the threshold that minimizes the intra-class variance. If the image intensity *f(x,y)* less than the threshold T this associated with the background, otherwise it is associated with the object (see Figure 4.7). Thresholding defines a mapping function M of the input image *f(x,y)* and producing the thresholded output *T(x,y)* so that [51]:

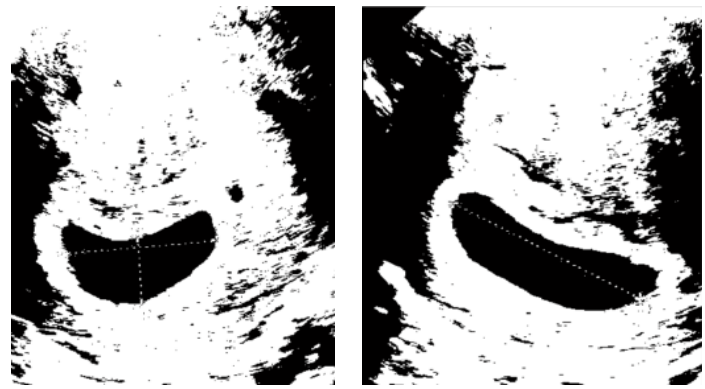
$$T(x,y) = M(f(x,y)) \tag{4.2}$$

In our work, we first applied the Otsu thresholding on the enhanced image to obtain a binary image as shown in Figure 4.7(a). In the binary image, the GSs as well as a number of false regions and small irrelevant objects are isolated from the background.

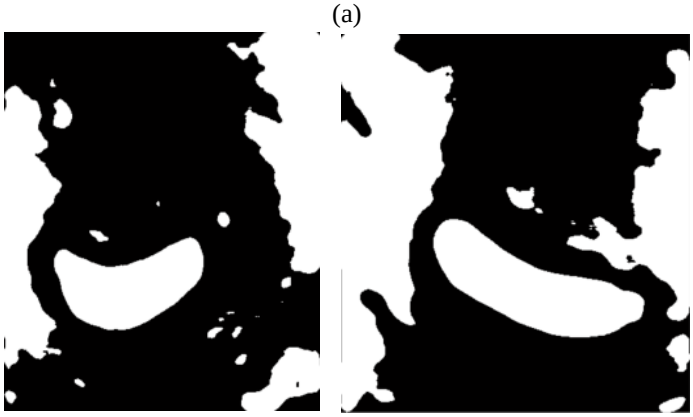
Step 2: Smoothing. A median filter (Gonzalez, Woods et al. 2004) with a window size of 15x15 is applied to the resulting image of step 1. As illustrated in Figure 4.7(b). This operation smooths the boundary of the sac without losing its original shape, fills small holes/gaps in the GS region, and helps to connect the non-sac or false regions to image borders for later removal.

Step 3: Removing attached regions with border. The imclearborder function in Matlab is then applied to clear all false regions that are connected to the image border, resulting in a clean image well inside the frame as shown in Figure 4.7 (c).

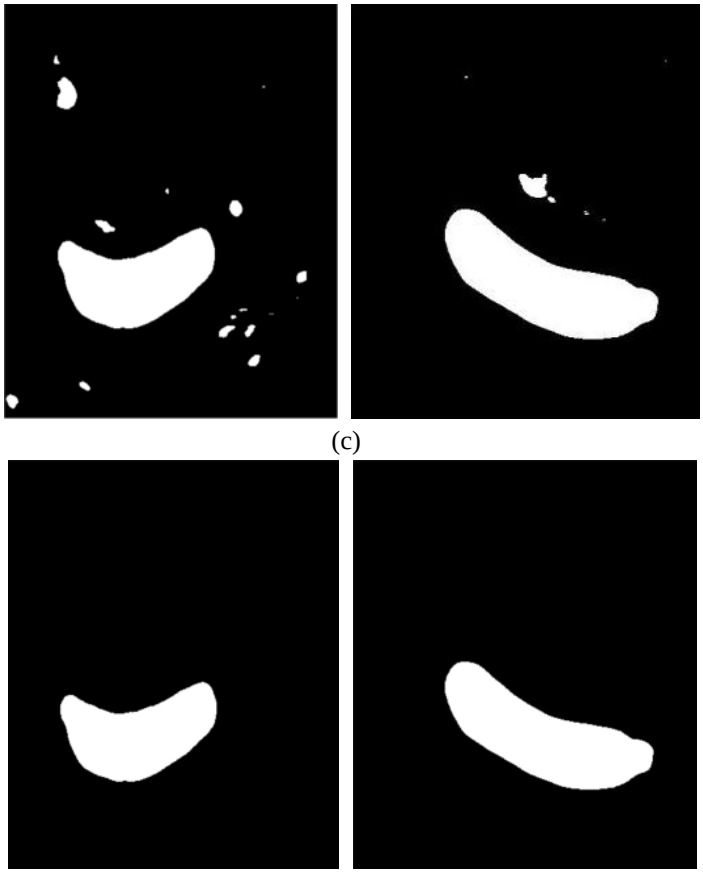
Step 4: False regions removal. Any small objects remaining in the image are considered as noises and should be removed. This was done by labelling each object using the Matlab function bwlabel, calculating the area of each object, and then deleting all small objects. The only remaining object is the GS as shown in Figure 4.7 (d). Our automatic segmentation was highly successful.



(a)



(b)



(c)

(d)

Figure 4. 7: Steps of GS segmentation (a) Binary image (b) Filtered image (c) Cleaned image from false region (d) Cleaned image from small objects

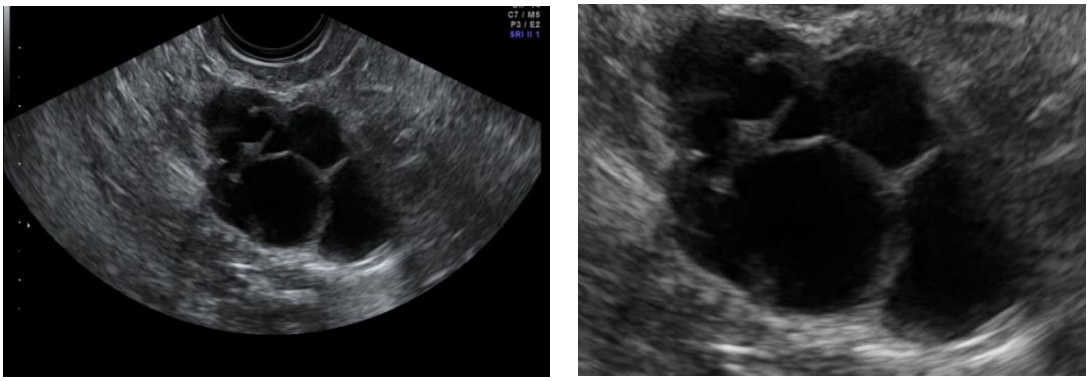
4.3 Enhancement and Segmentation of US Images of Ovarian Tumours

US images of ovarian tumours are usually corrupted with noise and unlike the GS, the region(s) of interest is complex and are difficult to distinguish from the surrounding by non-expert radiologists. This section introduces our automatic denoising and enhancing method for segmenting ultrasound images of ovarian tumours. We shall reoprts on the performance of our methods and highlight the limitations of the automatic segmentation of ovarian tumour.

4.3.1 Image Preparation

The margin areas of US images of ovarian tumour do not contain the region of interest, but affects the de-nosing and enhancing procedures and adds extra tasks that complicate the feature extraction stage. Therefore, we need to focus on the ROI and remove all unwanted areas such as the margin area and also the part of fan area which is considered as a background of the ROI. Accordingly we need to crop the image before applying the necessary enhancement and segmentation procedures. The experimental dataset are taken from a different machine and the tumour area is of variable size, and our algorithm would

have variable sets of parameters depending on whether the tumour is small or large. The imcrop Matlab function with a fixed position vector parameter may need to be modified through manual intervention. Figure 4.8 showed the example of the original image and cropped ROI. An example of cropped ROI is shown in Figure 4.8(b).



(a) The original image (b) The extracted sub-image

Figure 4. 8: An example of original image and cropped the ROI of ovarian tumour (a) The original image (b) The extracted sub-image

4.3.2 De-Noising and Enhancement Procedures:

Ultrasound scan can perform better than tumour markers or other medical image modalities in diagnosis of ovarian tumour (Timmerman, Schwärzler et al. 1999, Timmerman, Testa et al. 2005, Van Calster, Timmerman et al. 2007, Timmerman, Testa et al. 2008, Timmerman, Van Calster et al. 2010, Van Calster, Van Hoorde et al. 2015), de-noising and enhancing the ultrasound images of ovarian tumour helps the experts to improve their diagnosis by highlighting the anatomical information. The general requirements for removing the artefact in the ultrasound image are to suppress speckle noise as much as possible without affecting the anatomical information. The Non-local means filter is our choice of noise removal mechanism, inspired by the work in (Buades, Coll et al. 2005).

The Non-local Means Filter (NL-means)

Images are full of redundancies and many small blocks exhibit a great deal of similarity. The Non-local means filter, proposed by (Buades, Coll et al. 2005) differs from traditional filters in that it attempts to exploit redundancy and self-similarity in the input US image, and hence enables the removal of noise while keeping almost all meaningful information algorithm which will be helpful in area of medical image de-nosing (Buades, Coll et al. 2004, Buades, Coll et al. 2005).

Basically, the NL-means filter is based on non-local averaging of all pixels in the image that have similar neighbourhoods instead of taking the average in the filter neighbourhood. It estimates each pixel's intensity from the information provided from the entire image and hence it exploits the redundancy caused due to the presence of similar patterns and features in the image. In this technique, the returned grey value of each pixel is found by the weighted average of the grey values of all pixels in the image but with different weights. The weight assigned to each pixel is dependent on the level of similarity between the local neighbourhood of the pixel under consideration and the neighbourhood corresponding to other pixels in the image. An example of self-similarity is illustrated in Figure 4.9, below, which shows three pixels p, q1, and q2 and their respective neighbourhoods. The neighbourhoods of pixels p and q1 are similar, but the neighbourhoods of pixels p and q2 are not. Because of this, pixel q1 will have a stronger influence on the de-noised value of p and q2.More details can found in (Buades, Coll et al. 2004, Buades, Coll et al. 2005). Figure 4.10 shows an example of a US image obtained from a scan of an ovarian tumour, together with its NL-mean filtered version.

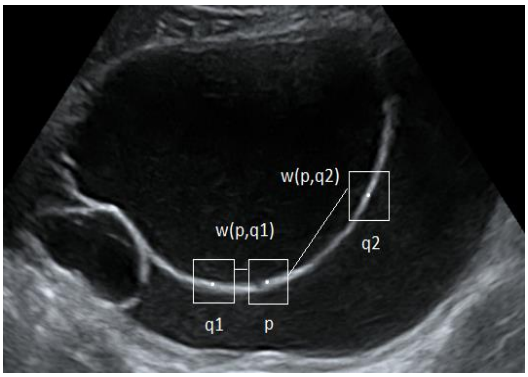
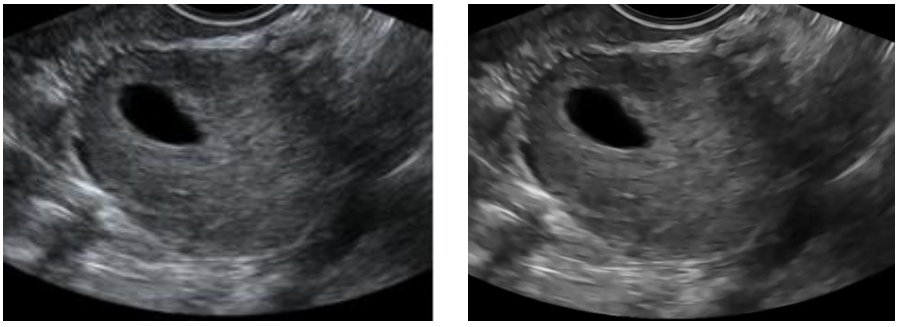


Figure 4. 9: Represents the blocks of neighbourhood pixel similarities



(a) (b)

Figure 4. 10: An example of a typical noisy US image and the NL-means filtered version Original Image (b) De-noised image

The Full Enhancement Method

De-noising is the major component of our enhancement scheme which consists of two more steps as illustrated in Figure 4.11. The NL-means filter is only the first step, while the full enhancement by taking the absolute difference between the NL- means filtered image and its negative.

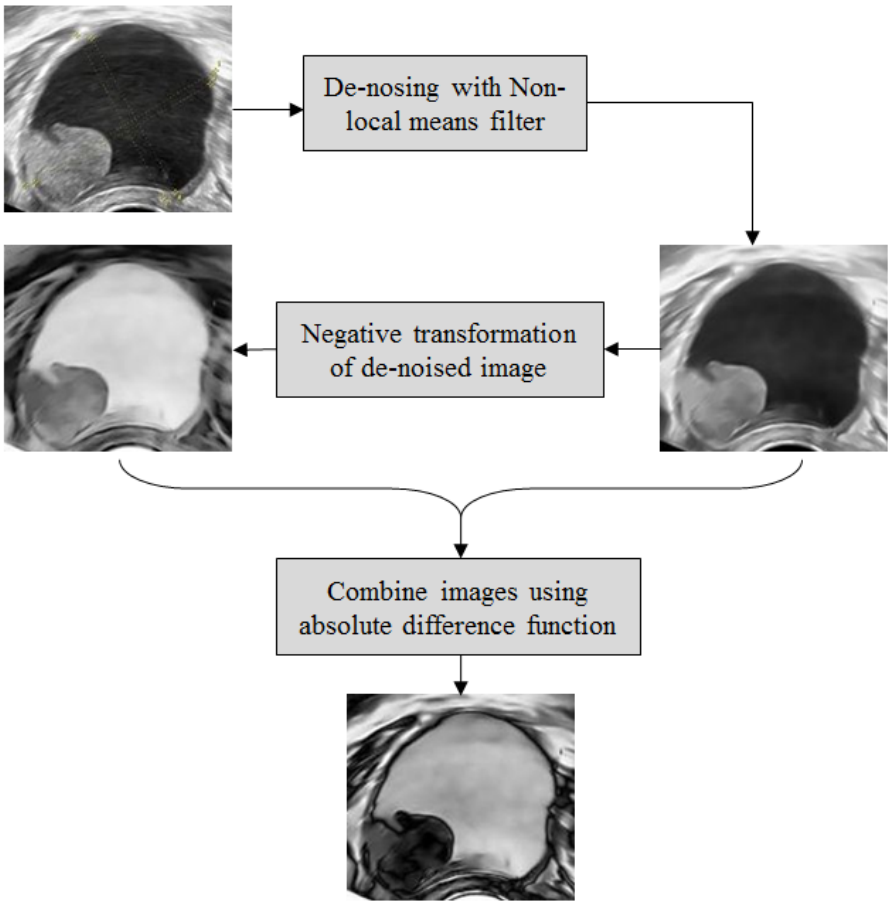
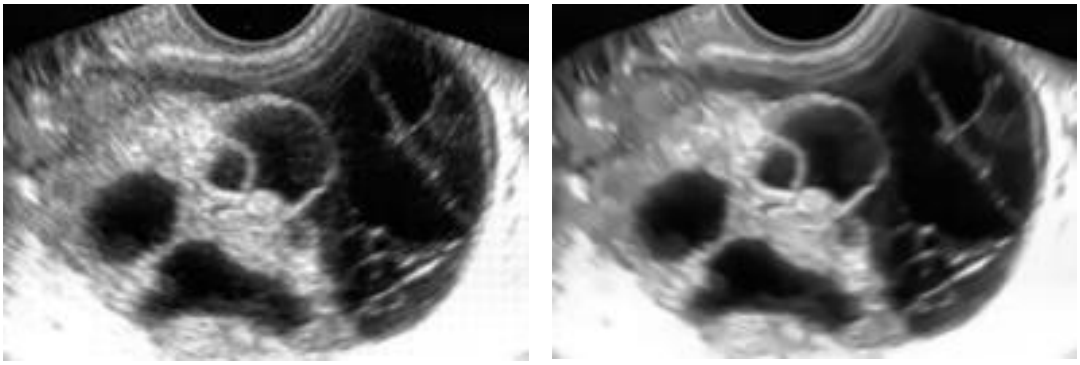


Figure 4. 11: Block diagram of the enhancing ultrasound image of ovarian tumour

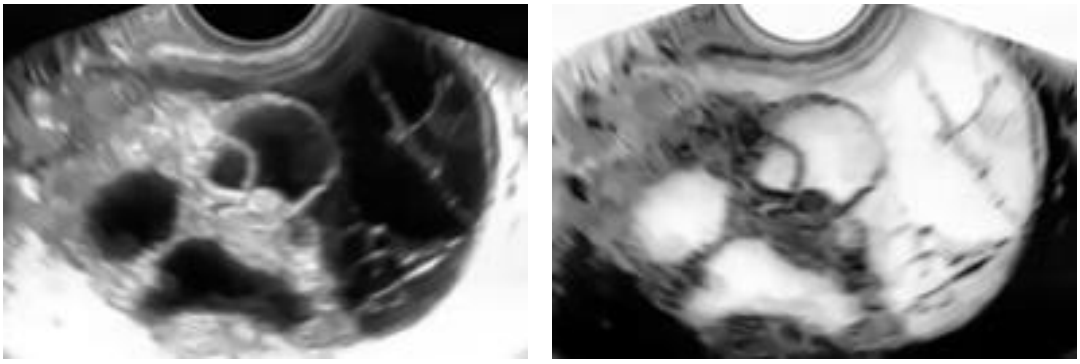
Step 1: Apply the NL- means filter on the input image to de-noise it, with kernel ratio (radius of local Patch) =3, window ratio (radius of neighbourhood search window) =3, and filter strength (strength of the NLMF filtering = 0.05). Figure 4.12 shows an example of an original image and its de-noised image.



(a) (b)

Figure 4. 12: Effects of NL-means filter (a) original image (b) De-noised version

Step 2: Compute the Negative of the de-noised image output in step 1, to enhance white or grey detail embedded in dark regions of an image (Gonzalez, Woods et al. 2004). Figure 4.13 is an example of the negative transformation of ultrasound images of ovarian tumour.



(a) (b)

Figure 4. 13: The negative of the de-noised image (a) The de-noised image (b) It negative image

Step 3: Finally, in this step we compute the absolute difference function between de-noised image and its negative transformed version. The Absolute Difference between the corresponding pixel values I1(i) and I2(i), from the two input images I1 (the filtered image) and I2 (the negative of the filtered image) is defined as:

$$r(i) = |I1(i) - I2(i)| \tag{4.3}$$

Where r(i) is the i^{th} pixel value in the resulting image. We apply the absolute difference operation on the de-noised image from the NL-means filtering step and its negative image. This means that

$$r(i) = |Intensity_{max} - 2 \times I(i)| \tag{4.4}$$

This means not only that the resulting image is negative of the original scan image, but also that the edges which are marked as white or light grey pixel color will become darker in the resulting image, thereby preserving and highlighting the edges. An example is presented in Figure 4.14, and the result image clearly shows the highlighted edges and the texture inside the tumors are much clear (the light grey shades in the resulting image).

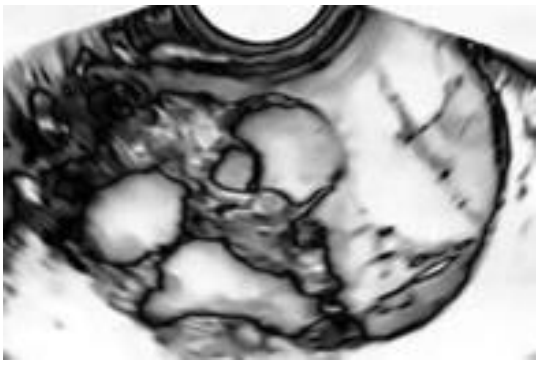
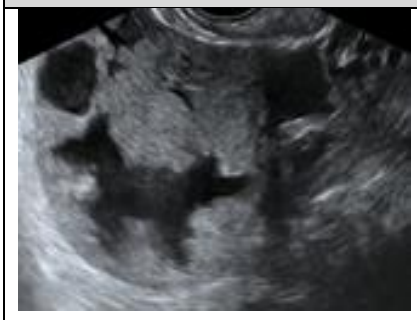
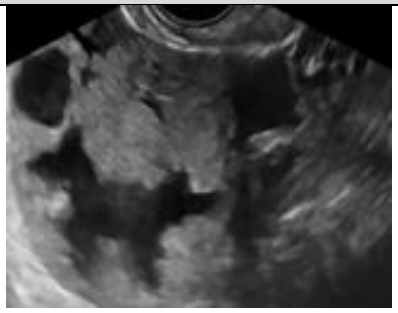
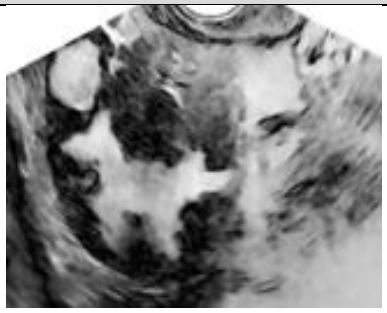
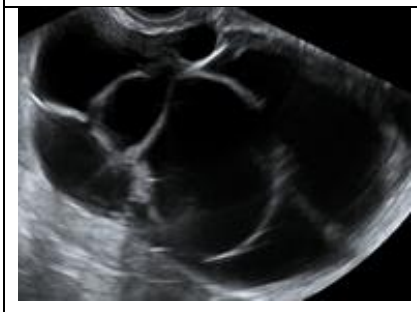
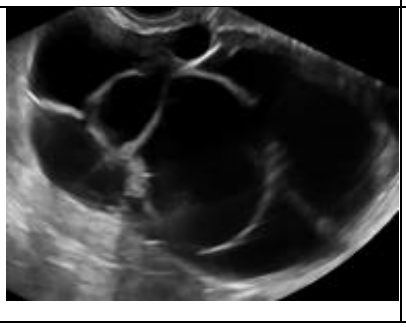
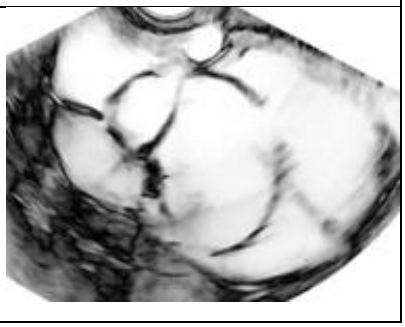
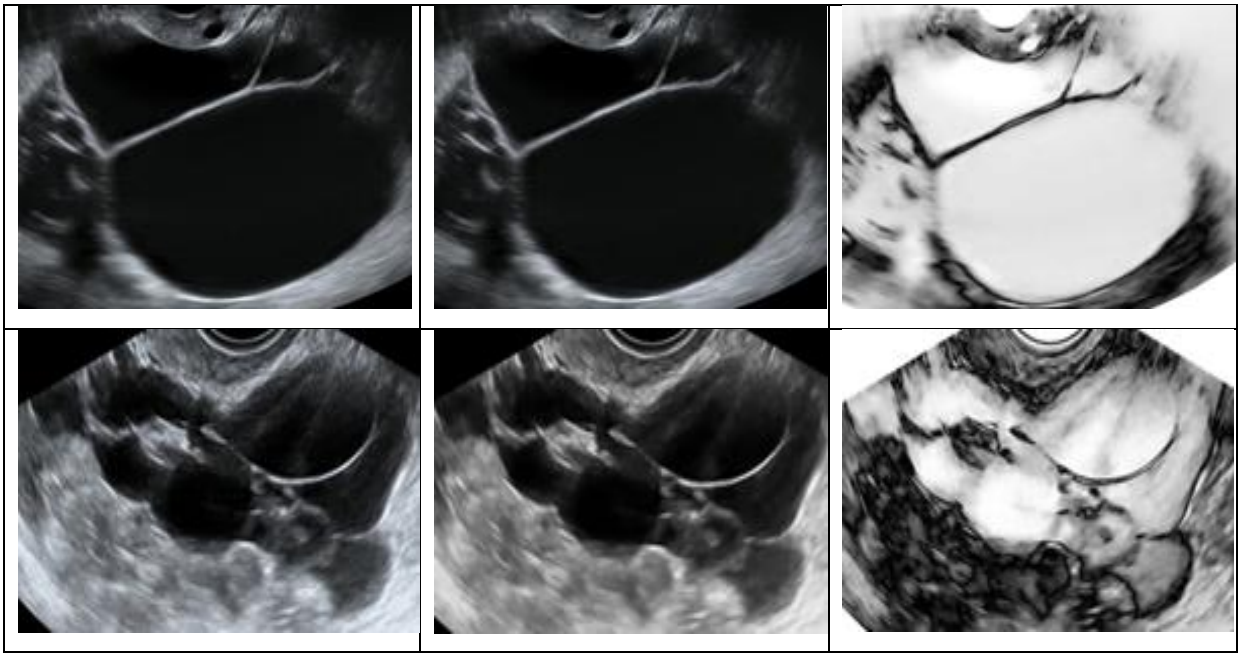


Figure 4. 14: The enhanced image all inner border are clear which is preferred by experts

More examples of the ultrasound image of ovarian tumours denoised and the result of enhancing are shown below in Table 4.1 column (1) is original images, column (2) is denoised images with NL-means filter and column (3) is the last result of enhanced images.

Table 4. 1: Example of de-noised and enhanced images of ovarian tumours

Original image	De-noised images	Enhanced images
		
		



Non-medical experts, like myself, may notice that the output of the three steps enhancement procedure are clearer images, but such a naive comment may hide obvious ignorance of the medical significance of the procedure. We evaluated the result of this three-step enhancement of the ultrasound images of the ovarian tumour by sending our dataset of US images of ovarian tumour, before and after enhancement, to three domain experts at Queen Charlotte's and Chelsea Hospital. The feedback we received was as follows: "*The enhanced images were found to be helpful. the results are interesting in highlighting the borders particularly in the irregular border of the tumours. This improves our assessment for these tumours. On another scale, it highlights the difference between echogenicity in irregular malignant tumours".*

Table 4.2 displays three different diagnostic results for a sample of cases: before enhancement, after enhancement, and actual diagnostic which was based on follow-up surgery/ biopsy.

For example, image 24, when the original image was used without enhancing, the result of the subjective assessment to diagnose this specific case was "Benign", with the degree of certainty of "Certain Benign". However, when used the enhanced image instead, they diagnosed the case associated with this US imageas "Malignant" with the degree of certainty of "Possible Malignant". And at the end the actual "Malignant" diagnosis was provided from the original source. This indicates that our enhancing method helps to understand better the images.

Table 4. 2: US images of ovarian tumour: Before, After and the Actual Diagnostics

Image ID	Diagnosis before Enhanemnet	Degree of Certaintly	Diagnosis after Enhancment	Degree of Certaintly	Actual Diagnosis
1	Benign	Certain benign	Benign	Certain benign	Benign
9	Malignant	Possible malignant	Malignant	Possible malignant	Benign
24	Benign	Certain benign	Malignant	Possible malignant	Malignant
30	Malignant	Possible malignant	Malignant	Certain malignant	Malignant
40	Benign	Certain benign	Malignant	Possible malignant	Malignant
44	Benign	Possible benign	Malignant	Possible malignant	Benign
51	Benign	Possible benign	Malignant	Possible malignant	Malignant
52	Benign	Possible benign	Malignant	Certain malignant	Malignant
53	Malignant	Possible malignant	Malignant	Certain malignant	Malignant

Figure 4.15 summarises the following findings from the table above:

1. No change to diagnosis (50% of our dataset): when the diagnoses before and after image enhancement agree on the same diagnosis, of which 45% was correct diagnosis (i.e. it matches the actual diagnosis) e.g. image number 1 from the table above and 5% was incorrect diagnosis e.g. image number 9.
2. Enhancement has led to correct diagnosis (20% of our dataset) i.e. the decision was corrected to match the actual diagnosis after image enhancement e.g. the image number 40.
3. Enhancement has led to improving the diagnosis confidence (27% of our dataset) i.e. the correct decision has changed from possible to certain after the image enhancement e.g. image number 30.

4. Enhancement has led to incorrect diagnosis (3% of our dataset) i.e. the decision was correct before the enhancement and became incorrect afterwards e.g. the image number 44.

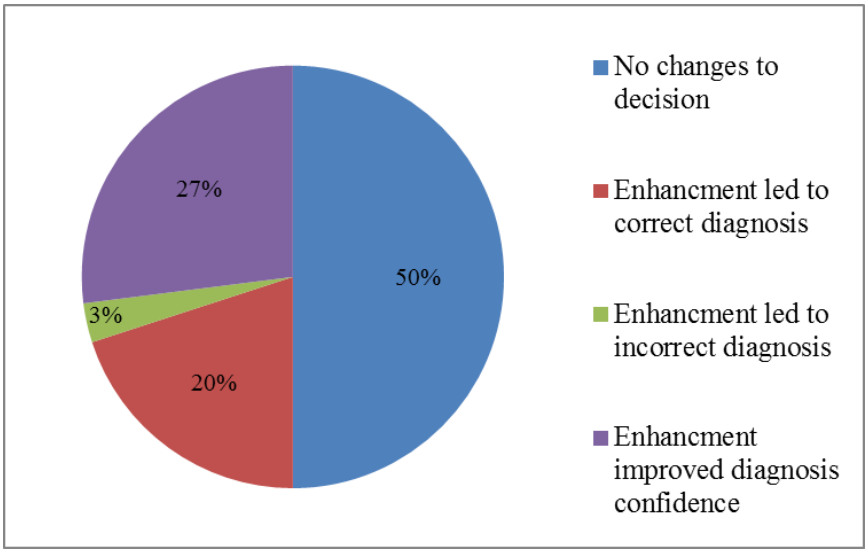


Figure 4. 15: Summary of the effect of the image enhancement on experts' decisions

A pilot enhancement software tool has been implemented together with a graphical user interface. The tool can be conveniently used by experts when they have a specific case that is not so easy to diagnose because of unclear information inside the tumour. This tool can help better understanding of the cases and thereby improve diagnosis accuracy. Figure 4.16 shows the screenshot of the interface of the tool.

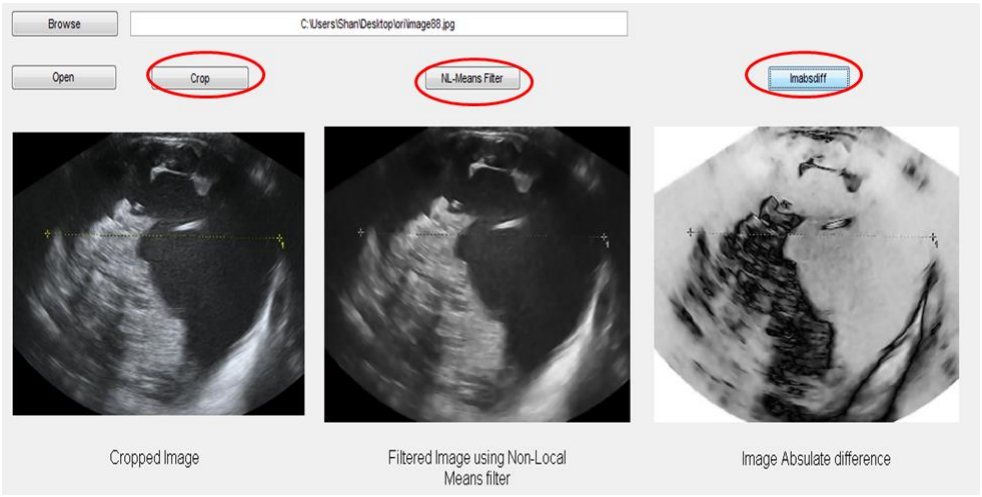


Figure 4. 16: Example of a Matlab Tool for End Users

4.3.3 A Semi-Automatic Segmentation of US Images of Ovarian Tumour

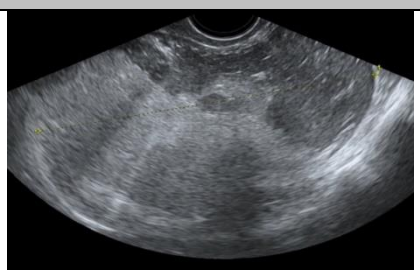
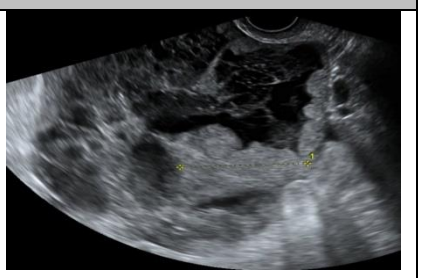
Enhancing the US scan images of the ovarian tumour is expected to enable the segmentation of the images and detect the tumour area from which digital diagnostics related features can be extracted from the spatial domain or the frequency domain. Though desirable, the automatic ovarian tumour segmentation was beyond the scope of this thesis. In fact, our main aim was to extract new image content features that are not usually considered by medical

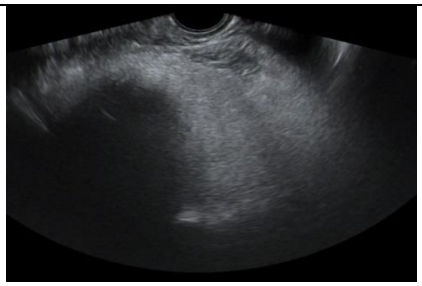
experts in tumour classification but may add extra support for experts' decisions. Moreover, so far automatic ovarian tumour segmentation is still an open problem and was reported in (Sohail, Rahman et al. 2010) (Hamid 2011) as a difficult challenge. In fact, the huge variation in tumour shape, size and position noted in our dataset of US images are some of the limitations that make a fully automatic segmentation of the ovarian tumour very complex that would hinder the feature extraction. Below we list a number of reasons for opting to adopt a semi-automatic approach to tumour segmentation.

1. **The image quality** affects the segmentation because the border of the tumour is not clear.
2. **The size of the tumour** in the images of our dataset can be very large and there can be an undefined boundary of the tumour inside the fan area of the ultrasound images.
3. **The similarity of texture** may exist inside the tumour area and its nearby background in some images. This problem leads to difficulty in automatically defining the border of the tumour especially when we have only one static B-mode image.
4. **The different textures may exist inside one tumour** in some cases. Besides, multiple components with their own inner borders, such as multilocular cysts, may exist inside one tumour.
5. **The combination of the above** may occur in a single image.

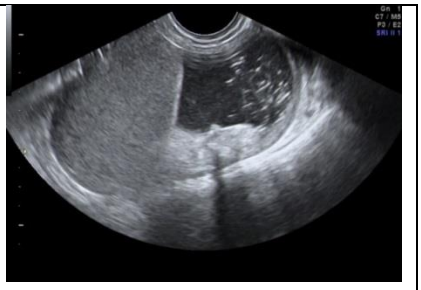
Table 4.3 lists some examples that illustrate the different problems mentioned above

Table 4. 3: Highlighting some examples that made automatic segmentation difficult

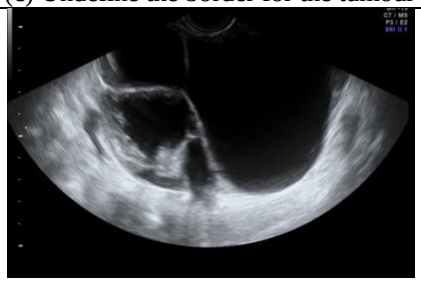
	
(a) Ovarian tumour texture and the background are very similar and missing the border	(b) Large ovarian tumour undefine the border



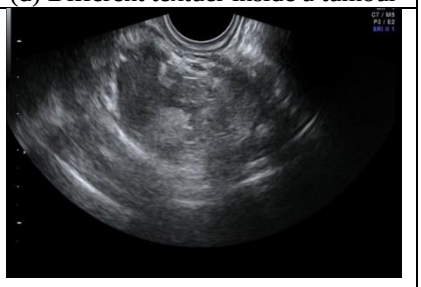
(c) Undefine the border for the tumour



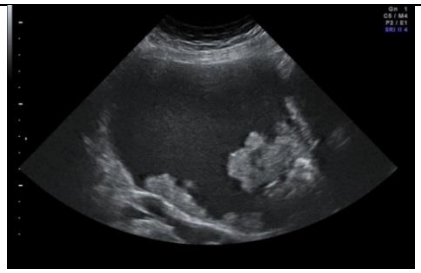
(d) Different textuer inside a tumour



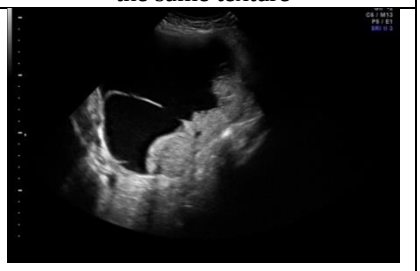
(e) Large size of the tumour



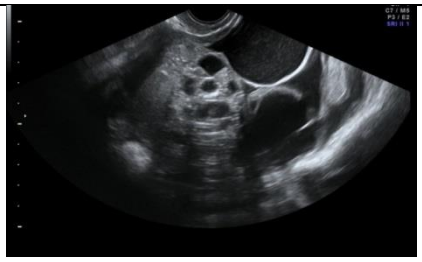
(f) Tumour and the background are the same texture



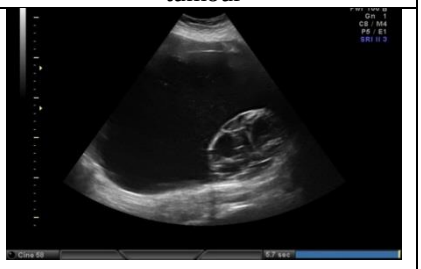
(g) Different texture representataion in the tumour



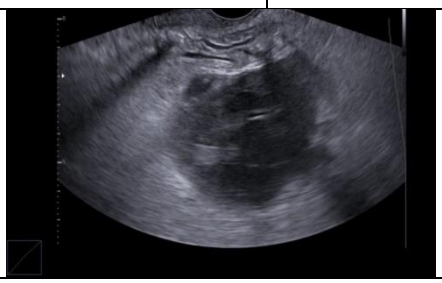
(h) Different texture in the tumour and undefined the border of the tumour



(i) Different texture and the large size of the tumour



(j) Large size of tumour and different texture

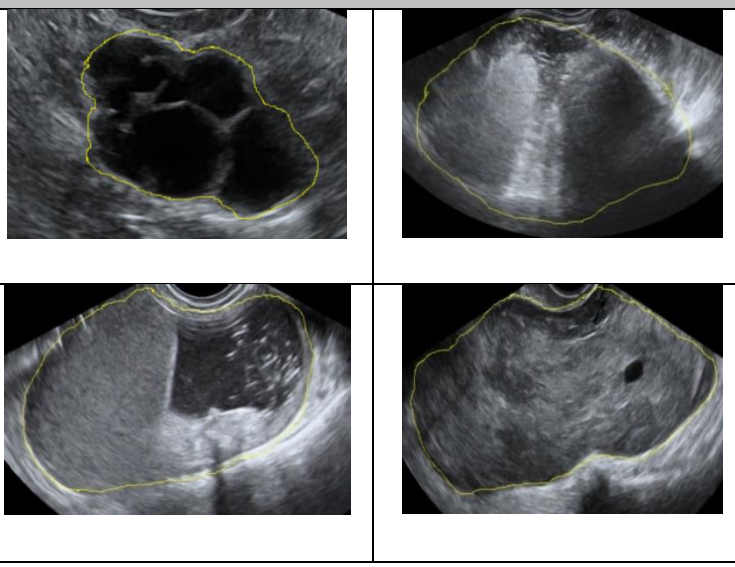


(k) Poor quality of the image

Despite the above observations, we are confident that automatic segmentation is still doable, but our interest was more on the reliability of feature extraction. The segmentation task was

therefore performed manually with guidance from domain experts, and the results have been reviewed and double checked by the experts. The manual segmentation was done using ImageJ that is widely used in biomedical image analysis. Table 4.4 presents some examples of the manually segmented ovarian tumour images defined by the yellow borders.

Table 4. 4: Example of highlighting the border of the ovarian using image J tool



4.4 Summary

In this chapter, we presented a literature review of the current existing work on ultrasound image de-noising and enhancement as well as segmenting the region of interest. We argued that there is no single method or algorithm that can effectively work for enhancing US images from all kinds of problem domains and extracting the regions of interest from the enhanced images. Moreover, we highlighted the fact that the ovarian tumour segmentation is still an open research problem. Furthermore, we introduced our method for GS extraction and enhanced the image with promising results. We have shown some images that were unsuccessfully segmented.

We also described in details our method for enhancing ultrasound images of ovarian tumour. We showed that the enhancement method helped to better diagnosis by highlighting hidden information in the image. The positive feedback from domain experts indicates that the enhanced images do help the experts in their diagnosis of the tumour status. The effect of enhanced images in term of feature extraction and automatic classification will be discussed in Chapters 5 and 6. Furthermore in this chapter, we highlighted the main difficulties that

automatic ovarian tumour segmentation methods face, especially for complex and challenging images.

In the next chapter, we shall use the enhanced images and the segmented ROIs resulting from this chapter to extract the significant and meaningful features for accurate diagnosis. These features of interest would expand the set of features currently used by medical experts to include new image-content based features that could add more information and value for medical diagnosis.

CHAPTER 5

CHARACTERISING FEATURES OF ULTRASOUND IMAGES IN GYNAECOLOGY

Having pre-processed the US images and segmented the ROI for the gynaecological abnormalities under investigations, the next step in the analysis of these images is to extract relevant features that will form the basis for the analysis and diagnosis. In image-based automatic pattern recognition applications, the main goal of feature extraction is to obtain the most discriminating information from an image. Such features are either encapsulated in the image spatial domain and/or after transforming it into another domain such as frequency or wavelet domain. The extracted features are then represented as feature vectors to be fed into a classifier to the next stage in the identification process. The main investigations in this chapter are concerned with complementing existing features, currently determined manually in clinics for diagnosis, by innovative sets of spatial/frequency domain features that can be used in the analysis of gynaecological ultrasound images for both types of investigated abnormalities (i.e. ovarian tumour and miscarriage). We investigate, develop, and test the performance of new sets of image-related features in discriminating the different classes of the two investigated gynaecological abnormalities. The specific contributions in this chapter include:

- Investigating the effectiveness of local features in comparison to global features for US images.
- Proposing novel frequency domain features to be extracted from binarized FFT spectrum.
- Automating the quantification of known features of US images of the GS to identify miscarriage cases at early stages of pregnancy, and investigating and evaluating new geometric features of the GS for miscarriage detection.
- Investigating the contribution of image texture features in spatial/frequency domains of US images to discriminate different classes of abnormalities for both applications.

This chapter is organised as follows. In section 5.1, we review the literature about different feature extraction and/or texture analysis methods that used for analysing US images. Sections 5.2 describes our proposed feature extraction methods for US images of ovarian tumours in spatial domain and describe our novel method based on frequency domain

features, and demonstrate their discriminating power between benign and malignant tumours. Section 5.3 will be devoted to investigating the suitability of these and other combination of existing and new geometric features for the detection of miscarriage from ultrasound images of the gestational sac. In Section 5.4 we summarise the work in this chapter, present concluding remarks about the result and identify the next set of challenges.

5.1 Literature Review

5.1.1 Image Features and Features Extraction

Features are a set of data attributes that encapsulate useful information about an entity (i.e. image), and ideally, it should be discriminating attributes that can be used to distinguish one sample from another. Basically, the process of transforming an input sample data into a set of features (usually called feature vector) is referred to as feature extraction (Lee and Landgrebe 1993) and (Guyon and Elisseeff 2006). In image-based pattern recognition applications, features extraction is often preceded dimensionality reduction to deal with the curse of dimension problem, and image information can be extracted from different image-representation domains (Spatial or Frequency) (Nixon 2008). Moreover, these features could be derived from the texture analysis of images to represent information about the arrangement of colours and/or intensities in an image. The main goal of a feature extraction technique is to extract the most relevant and discriminating attributes from the original data and represent the extracted feature vector in a space of lower dimension. If the feature extraction is carefully chosen, the extracted feature set is expected to perform well at the classification stage. In general, extracting feature vectors that yield optimum classification results and work for all relevant images all the time is a serious challenge. We note that capturing images of the same object do not produce the same raw data due to a variety of factors that cannot be perfected. In a typical image analysis application, various image pre-processing techniques (i.e. de-noising, image enhancing and ROI extraction) are required before extracting useful features (Nixon 2008). In this chapter, we assume that the US images have been pre-processed, enhanced and segmented as required.

5.1.2 Image Texture and Texture Analysis

In general term, image texture is a quantitative property that is meant to describe the smoothness (or otherwise) of the surface geometry and structure of the object as it appears in the image. Therefore, in image processing texture is defined as an attribute representing the spatial arrangement of the grey levels of the pixels in an image or a specific region of interest.

The texture is one of the most common features used to analyse, and interpret medical images (Kurani, Xu et al. 2004). In the context of this thesis, texture is defined in terms of the spatial distribution and variation of pixel grey value (intensity) of B-mode images. In general, there are different methods to compute texture features.

The texture analysis is the term used for methods developed to quantify image texture and describe image properties by textural features (Mathias, Tofts et al. 1999) (Nailon 2010). The purpose of performing texture analysis is to define a set of texture features that will identify the important properties of a texture for a defined ROI (Smutek, Šára et al. 2003). A texture analysis of medical images is widely used to differentiate between pathological and healthy tissues or to the segmentation of specific anatomical structures (Castellano, Bonilha et al. 2004).

5.1.3 Texture Analysis Approaches

There are several ways to categorise different texture analysis approaches. However, the most common way is to categorise them into Statistical-based, Model-based, Structural-based, and Transform based texture analysis (Materka and Strzelecki 1998, Bharati, Liu et al. 2004, Castellano, Bonilha et al. 2004, Materka 2004).

5.1.3.1 Statistical-based Texture Analysis

Statistical parameters of the texture features distribution are the basis of early methods investigated in machine vision (Castellano, Bonilha et al. 2004). Due to their simplicity and the ease of implementations, these are widely used in medical image analysis including ultrasound images (Bharati, Liu et al. 2004). Statistical based approaches can be further sub-divided into three categories: first-order, second-order, and higher-order description statistics (Materka and Strzelecki 1998).

1. **First-order description statistics**: this approach deals with the intensity of the pixels and measures the likelihood of observing a grey value at the image. In other words, the occurrence probability of intensity in an image (the image histogram) is the main source (Gonzalez, Woods et al. 2004). It does not take into account the spatial relationship between pixels but depend only on individual pixel values (Materka and Strzelecki 1998). In this thesis, First-order description statistics methods are employed using histograms and their First order statistical parameters.
2. **Second-order description statistics:** this method is taking into account both the distribution of pixels intensity and the relationship among neighbouring pixels

(Materka and Strzelecki 1998). An example of this method is Grey level co-occurrence matrices (GLCM), which used to express the correlation of spatial location and grey level distribution of an image.

3. **Higher-order description statistics:** this method deals with the relationship between three or more pixels (Materka and Strzelecki 1998). Local Binary Pattern (LBP), which is a grey-value image operator that represents a pixel in terms of its order relation to the certain pattern of its immediate neighbours, can also be considered as higher order statistical features. We adopt LBP in our work, and more detail description is given later.

5.1.3.2 Model-based Texture Analysis

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 (Materka and Strzelecki 1998). An example of the model-based texture analysis approach is Markov random fields (MRF) (Kervrann 1995).

5.1.3.3 Transform-based Texture Analysis

Texture properties of the image can be analysed in different image domains other than the spatial domain. Transform-based (also known as signal processing) methods apply texture filtering for extracting features but after transforming the image into different transfer domain image representation (Materka and Strzelecki 1998). Examples of model-based texture analysis approach are Fourier, Gabor and wavelet transforms. In this thesis, we investigate and develop a new novel transform-based method using Fast Fourier Transformation FFT. More details are explained in (section 5.1.5.4 and 5.2.2).

5.1.3.4 Structural-based Texture Analysis

The texture is characterised by texture primitives or texture elements, and the spatial arrangements of these primitives. Thus, the primary goals of the structural approaches are firstly to extract texture primitives, and secondly to model or generalise the spatial placement rules. The texture primitives can be as simple as individual pixels or a region with uniform grey scales. The placement rules can be obtained through the modelling geometric relationship between primitives or learning their statistical properties. The advantage of the structural approach is that it captures the semantic of the image, in terms of its content, providing a good symbolic description of the image (Castellano, Bonilha et al. 2004).

It is obvious that these three different approaches to structural analysis can be combined in a variety of ways. For example, it is rather natural to use extract statistical texture features from frequency rather than spatial domain representation of images. In fact, all the statistical features can be extracted from wavelet transformed images. To some extent, the Local Binary Pattern (LBP) texture method is a combination of the structural and statistical approaches. In fact, the LBP has both of the properties of a structural analysis method: texture primitives and placement rules. On the other hand, the distribution is just a statistic of a non-linearly filtered image, clearly making the method a statistical one. For these reasons, it is to be assumed that the LBP distribution can be successfully used in recognising a wide variety of texture types, to which statistical and structural methods have conventionally been applied separately (Ojala, Pietikäinen et al. 1996, Materka and Strzelecki 1998, Ojala, Pietikäinen et al. 2002). The pattern recognition literature has ample examples of successful use of LBP in different application such as face recognition and medical images (Mäenpää and Pietikäinen 2005, Pietikäinen, Hadid et al. 2011).

5.1.4 A Review of Existing Texture Analysis Methods for US Images

In the medical domain, interesting results of using texture analysis methods were reported. In the area of ultrasound images, many studies have been carried out to characterise the B-Mode images. Texture analysis of ultrasound images is motivated by the principle that if disease processes affect the structure/texture of tissues, then the diseased tissue should reflect ultrasound wave signals in a different way than that of normal tissue (Morris 1988), i.e. texture features value extracted from the US scan of diseased tissue differs from those extracted from US scanned healthy/normal tissue. The transformation of cancerous tissue, for example, will result in the changes in the tissue characteristics. Therefore, it is expected that textural features derived from cancerous tissue and normal tissue will differ.

Texture analysis has been used in ultrasound image analysis for different applications. In (Malathi and Shanthi 2010), a method for automatic classification between ultrasound images of normal and abnormal placenta was reported. The textural feature was extracted from the ultrasound image of placentas based on statistical properties of intensity histogram such as mean, standard deviation, contrast, correlation, and entropy. Another study showed that grey level co-occurrence matrices (GLCM) features derived from ultrasound images have a significant difference between the placenta of smoker and non-smokers where fourteen features were extracted from the GLCM matrices and in four different angles 0°, 45°, 90° and

135° (Morris 1988). The extracted features were an angular second moment, contrast, correlation, inverse different moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, two information measures of correlation, autocorrelation, and absolute value.

Furthermore, a similar method based on the fourteen GLCM features in the four dimensions was proposed in the area of automatic detection of heart disease (Tsai and Kojima 2005), The features were extracted from ultrasound images to classify heart disease where four features out of the fourteen were identified to have the most powerful discrimination ability (angular Second moment, contrast, correlation and sum entropy).

The texture analysis has also been applied extensively on ultrasound images for liver related diseases. Statistical-based texture analysis (GLCM) matrix was utilised to evaluate the texture features of the liver tumours in (Xian 2010). Five textural features were extracted from the GLCM matrix (energy, contrast, correlation, entropy and homogeneity) using the average of the four directions. It was demonstrated that the method can distinguish between benign and malignant tumours of the liver. Moreover, in (Mojsilovic, Popović et al. 1998), wavelet transform features were used to capture texture characterisation of B-mode liver images where the energies of the transformed regions were used to characterise the textures.

Texture analysis of ultrasound images has been also applied on thyroid gland. A promising result was reported in (Smutek, Šára et al. 2003) in which chronic inflamed thyroid tissues were automatically differentiated from healthy ones. The extracted features were based on the following six spatial features: gradient magnitude, difference from sample mean, horizontal curvature, vertical curvature, and original pixel grey levels. The six features were then combined with the following nine haralick texture features: cluster tendency, texture entropy, texture contrast, texture correlation, texture homogeneity, inverse difference moment, maximum probability, and uniformity of energy.

Furthermore, texture analysis was used to discriminate between benign and malignant breast tumours from ultrasound images. In (Lefebvre, Meunier et al. 2000), the texture parameters, derived from first-order statistics, run-length matrices and co-occurrence matrices. It was argued that texture features were able to help physicians in reducing the number of unnecessary biopsies. In (Garra, Krasner et al. 1993), four different types of ultrasound images of breast tumours were studied namely, cancers, cysts, fibroadenoma, and fibrocystic tumours. The analysis of image texture was performed using fractal analysis and

statistical texture analysis methods. The most useful features were those derived from the GLCM matrices based on the contrast and correlation features. It was argued again that ultrasonic image texture analysis is a simple way to significantly reduce the number of benign lesion biopsies without missing additional cancers.

Besides breast cancer, texture analysis has been also applied on ultrasound images of prostate. In (Sheppard and Shih 2005) GLCM is generated based on the above fourteen extracted features from prostate ultrasound images. Then, five of the fourteen features were chosen namely, angular second moment, contrast, inverse difference moment, entropy, and sum entropy. These features were successful in distinguishing between normal and cancer tissues in prostate US images.

A rather limited research has been conducted to evaluate texture analysis on US images of different types of ovarian tumours. In (Sohail, Rahman et al. 2010), an automatic method was proposed to classify three different types of benign ovarian tumours namely; Simple Cyst (187-images), Endometrioma (154-images) and Teratoma (137-images) in total 478 images were used. The features were extracted from statistical texture using 64 features based histogram moments along with 56 features extracted from GLCM in four directions. An average of classification accuracy 86.90% was achieved to identify different types of benign tumours.

In (Acharya, Vinitha Sree et al. 2012), Higher Order Spectra (HOS) features were used for the characterisation of different types of US images of ovarian tumours namely, Benign and Malignant. A small dataset of 20 patients (10 benign and 10 malignant) was used to evaluate the method where each patient has 100 images i.e. 1000 benign and 1000 malignant ultrasound image in total. A classification accuracy of 95.1%, sensitivity of 92.5% and specificity of 97.7%, was achieved to identify different types of ovarian tumours.

More research on automatic identification of US images of ovarian tumour was reported in (Acharya, Mookiah et al. 2013) where three different types of features extracted from each ultrasound image of 20 patients (10 benign and 10 malignant). These features are: Hu's invariant moments features (invariant to object scale, position, and orientation), 2D Gabor wavelet features at six directions, and Yager's measure and Kapur's entropy to estimate the subtle variation in the pixel intensities. An average classification accuracy of 99.8 %, sensitivity of 99.2 % and specificity of 99.6 % was reported to identify different types of benign tumours.

To the best of our knowledge, there is no publication in the literature on using texture analysis in the area of gestational sac to identify early miscarriage cases.

5.1.5 Texture Analysis Methods used in this Study

This section discusses the methods that we employed in this thesis for texture analysis of the US images of both investigated gynaecological abnormalities. The investigated methods include, Histogram of grey scale, First order statistical features of histogram image, Local Binary Pattern LBP, Fast Fourier transformation FFT. By extracting each of these features, we produced a feature vector of certain dimension as representing the US scanned images that will be feed into one of the chosen classifiers (kNN or linear SVM).

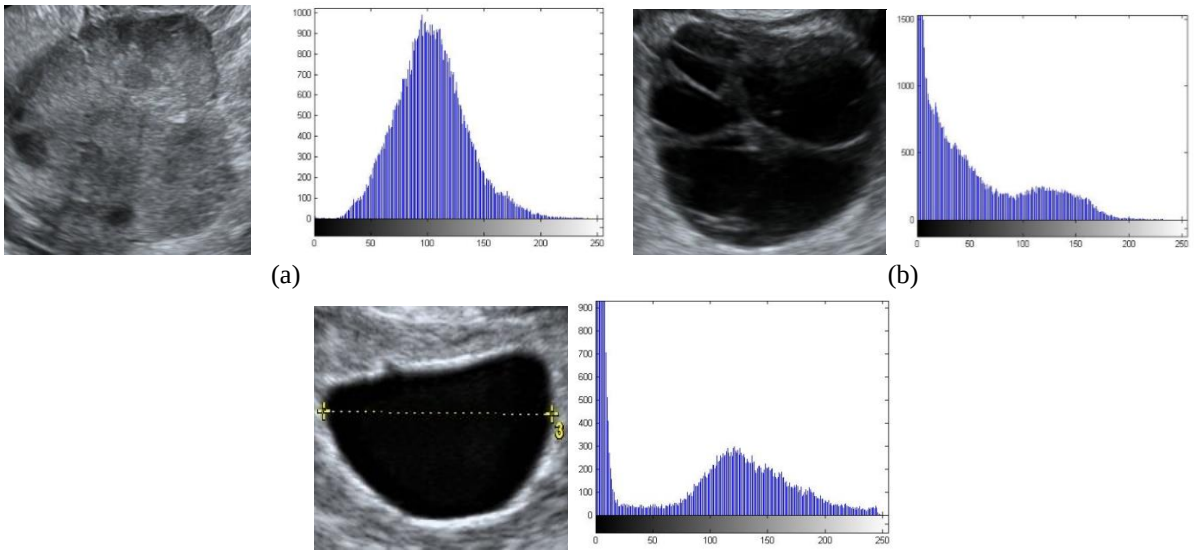
When using kNN classifier for automatic diagnosis of the two gynaecological abnormalities under investigation, we need to define a distance/similarity function between extracted histogram feature vectors extracted from the US scan images. Histogram intersection and Euclidean distance function are among the most popular comparison between histogram feature vectors of a template and a test image. For this purpose, we use the Euclidean distance function. But when using linear SVM, we shall use the inner product kernel function between the support vectors and the above-extracted features vector of the probe US images.

5.1.5.1 The Histogram of Grey Scale Intensity

Basically, an image histogram represents the frequency distribution of pixels intensity value in the image over the full grey-scale (Malathi and Shanthi 2010). For an 8-bit grayscale image there are 256 different possible intensity bins, and so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those grayscale values (Malathi and Shanthi 2010). The histogram of an image conveys information about brightness and contrast in the image. A comparison between the three histograms, shown in Figures 5.1 (a) - (c), reveals a significant difference which seems to provide a discriminating feature between an image of the malignant tumour (figure (a)) and those of a Benign one (figure (b)). The first histogram of the US scan of malignant tumour has a clear normal distribution most of the mass area is of grey colour, indicating the presence of solid tissues within the tumour. The histogram of the benign tumour has a negative exponential shape concentrated to the left side of the greyscale with a small hump around the middle of the greyscale, indicating the presence of clear fluid within the tumour. The third image is a scan of the gestational sac, and its histogram has two almost separated regions where the background is represented by a more of a normal curve near the middle of the dynamic range

and the inside of the GS to the left side of the greyscale because there are a dark region inside the gestational sac.

Histograms can also be used to study the distribution of various components of the images. For instance, it can be used to segment the ROI or distinguishing between two different classes if the two classes have different histogram distribution. But the main disadvantage is because it represents the frequency of the pixel intensity values but with no spatial information. Sometimes two scan images of different disease classes may have a similar histogram but they may be different in the content and texture.



(a) (b)

(c)

Figure 5. 1: Histogram distributions for different ultrasound images (a) US image of malignant ovarian tumour with its histogram (b) US image of benign ovarian tumour with its histogram (c) US image of uterus include gestational sac in a sagittal plane with its histogram

The Histogram feature vector extracted from a US image is the 256-dimensional vector:

$$\underline{v} = [v_0, v_1, \dots, v_i, \dots, v_{255}],$$

Where v_i is the frequency of the pixel value i in the image.

5.1.5.2 The First-Order (FO) Statistics Histogram Properties

Different useful statistical parameters can be worked out from the histogram to quantitatively describe the first-order statistical properties of the image. Most often the so-called central moments are derived from it to characterise the texture these features include mean, variance, smoothness, third moment, kurtosis and entropy (Levine 1985, Materka and Strzelecki 1998).

If z is a random variable representing an image pixel intensity, $p(z)$ is the histogram of the intensity levels in an image, and L is the total number of pixel in the image then the FO features can be computed using the equations (5.1 – 5.6), below:

Mean: The mean takes of the average level of intensity of the image, and show the brightness of the image.

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \tag{5.1}$$

Standard Deviation STD: The STD describes the variation of intensity around the mean (i.e. contrast).

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 \, p(z_i)} \tag{5.2}$$

Smoothness: The smoothness measures the relative smoothness of the intensity.

$$R = 1 - \frac{1}{1+\sigma^2} \tag{5.3}$$

Skewness: the skewness or (Third-moment) measures how the histogram is balanced around its mean and which side has the longer tail. The skewness is zero if the histogram is symmetrical about the mean. Otherwise, it is either positive or negative depending whether it has been skewed above or below the mean.

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i) \tag{5.4}$$

Kurtosis: The kurtosis measure how sharpness of the peak of a frequency-distribution curves.

$$U = \sum_{i=0}^{L-1} p^2(z_i) \tag{5.5}$$

Entropy: The entropy measure of randomness of the different intensity frequencies.

$$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{5.6}$$

Together all these features provide information about the shape and distribution of the histogram, such measurements might have a high discriminative power to distinguish

between different kinds of images (Castellano, Bonilha et al. 2004). The extracted feature vector from a US scanned image is a six-dimensional vector whose entries are the above six values in the above order.

5.1.5.3 The Local Binary Pattern (LBP)

There are many ways of describing the texture in an image, but in all cases, it refers to local changes in intensity. The LBP as a texture feature was first proposed by (Ojala, Pietikäinen et al. 1996) that refers/maps each pixel to an 8-bit binary code that is derived from the intensity of its neighbouring pixels within a given window. For a 3×3 block, the value of the centre pixel is subtracted from that of each of its eight neighbouring pixels. Depending on the sign of the subtraction result, the surrounding neighbouring pixels *s(x)* are set to 1 (for +) and 0 (for -) as shown in the equation 5.7. The generated bits of all the neighbouring pixels are then concatenated into 8-bit codes, starting with the top left corner neighbour and moving in a clockwise direction. The decimal value of the LBP code for the centre pixel (x_c, y_c) is calculated as in equation (5.8) where *n* runs over the eight neighbours of the central pixel, i_c and i_n are respectively grayscale values of the central pixel,

$$s(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{5.7}$$

$$LBP(x_c, y_c) = \sum_{n=0}^{n=7} s(i_n - i_c)2^n \tag{5.8}$$

The above definition can be generalised in different ways, depending on the size of the neighbouring window and the greyscale resolution. The LBP operator $LBP_{P,R}$ produces 2^p different output values, corresponding to 2^p different binary patterns formed by the P pixels in the neighbourhood of radius R as illustrated in Figure 5.2. Figure 5.3 shows the centre of an LBP block with different radius P=8, R=1(3x3), 2 (5x5).

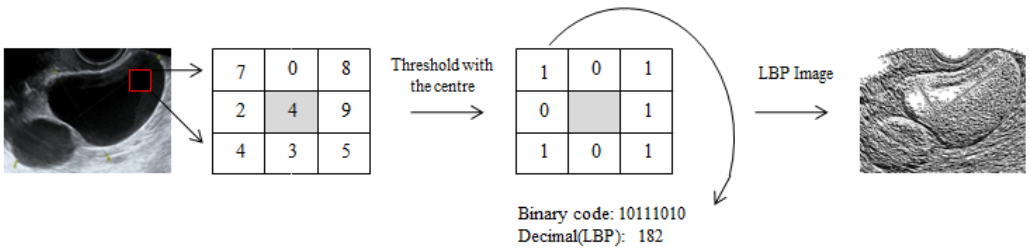


Figure 5. 2: Shows the LBP coding and the result of the LBP image with R=1

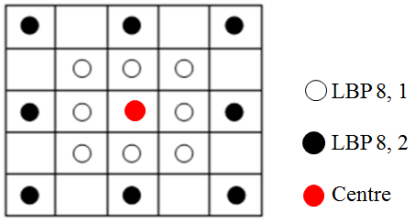


Figure 5. 3: Represent the centre with 8 neighbouring

One of the advantages of the image information provided by LBP is its invariance against monotonic grey level changes. Moreover, the most interesting property is its computational simplicity. The histogram of the LBP provides various texture discriminative methods, depending on the grouping of the different LBP patterns depending on a defining a similarity relation among the patterns. Grouping the different patterns results in creating a histogram bin for each group, so that the frequency of a bin is the total number of appearances of all the 8-bit LBP patterns in that bin. The most common types of LBP methods are:

1. **Simple LBP (256) bins:**

In this case similarity of LBP is defined as equality of the patterns. Therefore, the feature vector extracted in this case is simply the frequency of all the different 8-bit binary codes defined by the LBP mapping, i.e. it is the histogram of the LBP mapped image obtained from the different decimal grey-level values representing the 8-bit codes. In this basic type of LBP method, one is simply using the *256 bins* histogram of LBP map of the image.

2. **Uniform LBP (59 bins):**

The method is based on defining similarity between local binary patterns that is defined in terms of the number of transitions between 0 and 1 in the patterns. The most common approach in this case creates different classes of patterns one of which is termed as the "uniform" pattern, denoted $LBP_{P,R}^{u2}$, where P are the pixels in the neighbourhood of radius R. A local binary pattern is called *uniform* if it contains a maximum of two bitwise transitions from 0 to 1 or vice versa where the corresponding bit string is considered circular. For instance, 00000000 (0 transitions) and 01111110 (2 transitions) are uniform whereas 11101101 (4 transitions) and 01011011 (6 transitions) are not. It has been shown, that there are at most 58 different uniform 8-bit LBP patterns that occur in any image. Moreover, the distinct uniform patterns are deemed to encapsulate the most discriminating the distribution of texture features in any image. Therefore, uniform LBP histogram based method group together all the binary patterns that have more than two transitions (i.e. a non-uniform patterns) the total frequency of which will be under the 59th bin of the histogram, and the

frequency of the remaining uniform patterns will recorded in 1-58 bins of the histogram (Ahonen, Hadid et al. 2004).

Depending on the version of LBP scheme, the extracted feature vector is either of the form

$$\underline{v} = [v_0, v_1, \dots, v_i, \dots, v_{255}], \qquad for\ the\ simple\ LBP\ scheme, and$$

$$\underline{v} = [v_0, v_1, \dots, v_i, \dots, v_{59}], \qquad for\ the\ uniform\ LBP\ scheme.$$

Where v_i is the frequency of the pixel value *i-th* bin in the image, and in the case of uniform LBP the 59^{th} bin represents the frequency of the non- uniform patterns.

5.1.5.4 The Fast Fourier Transformation (FFT)

All the methods explained above were extracted from the spatial domain where an image is a matrix of grey-level intensity. The frequency domain of an image refers to the rate at which the grey-level intensity changes, providing a new representation of the image. In the case of images captured by a digital camera, the frequency representation of the image can be computed using mathematical transformations that map image pixel intensity with the frequencies of the "light" waveforms that hit the surface of the photographed scene. The same mathematical transforms can be applied to any other type of images, regardless of the capturing device and the generating source of waveforms, to obtain its frequency domain. In general, the frequency domain representation of an image captures particular changes in the image spatial domain, such as rapid changing in grey values which corresponds to high-frequency waveform (e.g. borders of image components). In contrast, the appearance of low frequency waveforms indicates minor changes in grey values which are associated with smooth regions. There are many mathematical transforms that help to model the frequency domain of images. The Fourier transform is the most common and the oldest form of mathematical tools that analyses an image into its spectral components determined by wavelength which is equivalent the waveform frequency (Gonzalez, Woods et al. 2004). Other useful transforms include discrete wavelets transforms and Discrete Cosine Transforms.

It is assumed that for any image waveforms of different frequencies contributes to the generation of a single pixel value. In fact, we already know that the Sun's visible light is a mix of a large number of different frequencies that form the colour of the rainbow, which can be seen when sun light passes through a prism. Naturally objects photographed reflect

different proportions of these contributing frequencies depending on texture and orientation of the object with respect to the light as well as the orientation of the object. The same apply to US images or images generated by any source of waveforms. The Fast Fourier Transform (FFT) is an efficient algorithm that returns the contribution of the different frequency waveforms that determine the pixel values of the image.

The Discrete Fourier Transformation DFT of an image extracts the strength of the different frequency waveforms contributing to the pixel values of the image (Gonzalez, Woods et al. 2004). The DFT signal and its inverse can be computed using FFT. The DFT of an image *f* for any frequency pair *(u,v)* is a complex number that depends on all the spatial pixel values *f(x,y)* computed by the formula:

$$F(u,v)=\left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cos(2\pi\,(u\,x/M+v\,y/N)\right. \\ \left.-\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\sin(2\pi\,(u\,x/M+v\,y/N)\right] \tag{5.9}$$

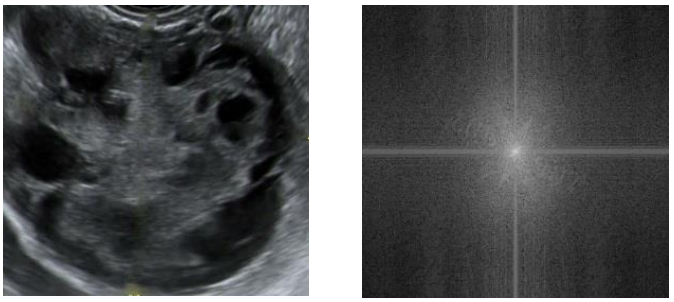
The output of the FFT is a size two array, which is also represented for each pair (u,v) of frequencies as a complex number whose real part Re(*F*) is the first entry in the above expression and imaginary part Im(*F*) entry. Unfortunately, we cannot display the output as an image. However, another more useful way of representing the complex numbers *F*(u,v) is in terms of the spectrum of *F* defined as its modulus:

$$\|F(u,v)\|=\sqrt{(\mathrm{Re}(F(u,v)))^2+(\mathrm{Im}(F(u,v)))^2} \tag{5.10}$$

and its phase:

$$\emptyset(F(u,v))=\arctan\left(\frac{\mathrm{Im}(F(u,v))}{\mathrm{Re}(F(u,v))}\right) \tag{5.11}$$

The spectrums of the FFT of an image have been used as an indicator of power possessed by the FFT. Figure 5.4 shows the ovarian ultrasound image with the spectrum of FFT.



(a) (b)

Figure 5. 4: Example of ultrasound image of ovarian tumour before and after FFT(a) Original image (b) FFT spectrum of the image

5.2 The Proposed Feature Extraction Method for Ovarian Tumours Classification from US Images

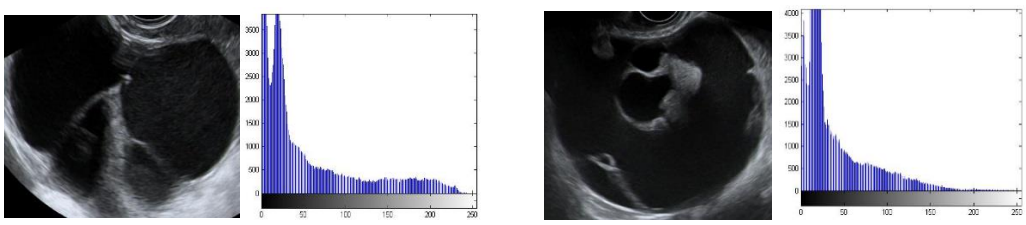
This section is concerned with the testing the performance of our proposed list of feature vectors defined in the last section. Apart from the FFGF feature vector model, all the other feature vectors can be extracted locally by subdividing images into blocks of different sizes. First we shall investigate the effectiveness of extracting those local features instead of the global ones by using blocked based version of the above feature extractions. We then propose test the performance of the FFGF feature vector extracted from ultrasound images in terms of its discriminating benign and malignant ovarian tumours.

In this section, all experimental results are based on ultrasound images that were cropped and pre-processed (*Enhanced ROI*) using the method that was described in (Chapter 4, Section 4.3). The aim of the current identification experiments is to rank the various feature vectors in terms of their discriminating power. We therefore only use the *linear SVM* classifier, described in (Chapter 3, Section 3.5) to evaluate and compare the performances of the different feature extraction schemes. In Chapter 6, we investigate the use of other classifiers. In these experiments, we adhere to the evaluation protocol that was explained in details in (Chapter 3, Section 3.7).

5.2.1 Extracted Local Features from the State-of-art Methods

5.2.1.1 Localised Histogram Intensities Features

Extracting feature vectors from the entire image (e.g. histograms) may not be ideal for discriminating two ultrasound images of different ovarian tumours. Figure 5.5 shows two images that have very similar histograms and different classes which are difficult separate by any classifier. However, there are obvious local variations in the intensity distribution, which is the motive for investigating the discriminating power of block-based local histogram feature vectors.



(a) (b)

Figure 5. 5: Two different types of ovarian tumours with their histogram (a) Benign tumour with its histogram (b) Malignant tumour with its histogram

We extracted local features instead of the global ones by dividing the images into sub-blocks, getting the histogram of each block, and concatenate all histograms to form a single feature vector representing the input image. Figure 5.6, illustrate a 3x3 blocks partitioning example.

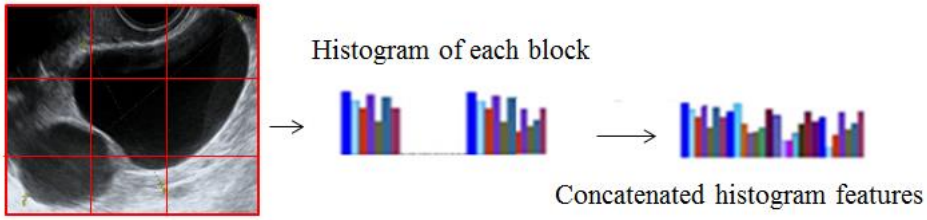


Figure 5. 6: an example of 3x3 blocks with concatenated histograms

We investigated the performance of 9 different sizes block subdivision. Figure 5.7 demonstrates the effectiveness of local features over the global one. The figure shows the accuracy rate obtained by the linear SVM classifier each one of these nine different subdivision schemes. We notice signification improvements in terms of classification accuracy from 66.46% for using the whole image to over 70% for most of the blocking schemes attempted and thereby confirm that local features can capture more information relevant to distinguishing malignant from benign ovarian tumours. The optimal accuracy of 76.33% is achieved for the 4x4 blocking scheme, but the figure also shows that using more blocks does not always help in capturing more relevant information. In fact, when we have 9x9 blocks, the extracted features might become too localised and thus, the overall accuracy deteriorates.

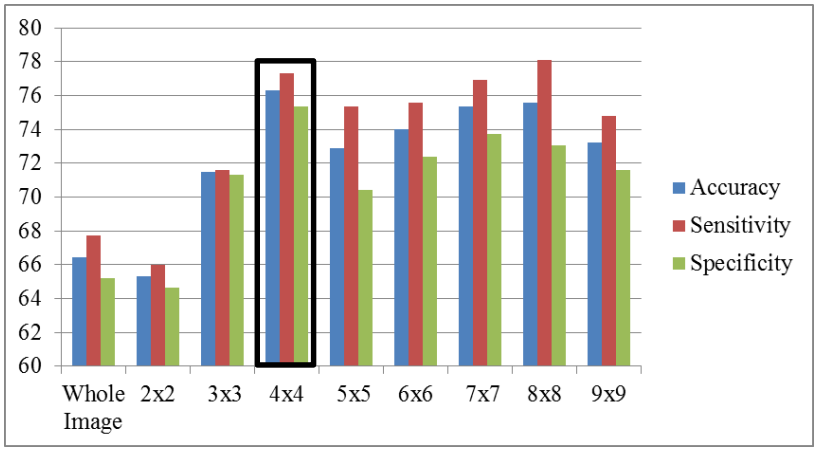


Figure 5. 7: Accuracy of local based histogram using SVM

5.2.1.2 Localised First-Order Statistical Features

As in the previous section, we extract local version of the six-dimensional statistical parameters of the US image intensity histogram for the same nine block-subdivision schemes. We calculated these features based on the equations presented in (Section 5.1.5.2). The corresponding feature vector of each block is six. For example, the 2x2 block subdivision scheme produces a 24-dimensional feature vector.

The performances of the various subdivision schemes as shown in Figure 5.8 confirm that the local features improved the accuracy of ovarian tumour classification as a result of capturing more relevant information to distinguish these two classes. Diagnosis accuracy was improved from 67.2% using a global feature to 73.46% and 74.46% using 4x4 and 5x5 blocks respectively. Again, the figure also shows that using more blocks does not help in distinguishing between different classes as features might become too localised and thus the overall accuracy decreases.

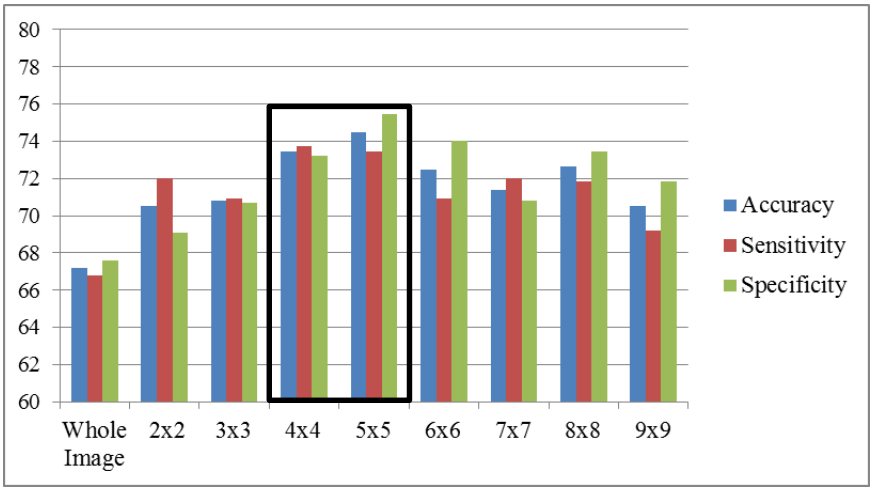


Figure 5. 8: Classification result based on local statistical histogram features using SVM classifier

Moreover, diagnosis accuracy of the statistical features of histograms (74.46%) is slightly lower than those based on using the whole histograms (76.33%) presented in the last section, indicating the limitations of the localised six moments in capturing local variations.

5.2.1.3 Localised Histogram of LBP Features

Having established that the localised 4x4 feature vectors have resulted in best/high performance when used with both previous feature vectors, we first conducted an experiment to compare the performances of the localised simple LBP histogram (i.e. the 256 bins scheme) with that of the uniform LBP histogram (59 bins scheme) when the images were

divided into 4x4 blocks. We also tested the effect of removing the 59th bin that contains all non-uniform patterns. We want to establish which type of LBP histogram has the optimal classification performance. The second experiment, in this section, is designed to test the performance of localised versions of the optimal LBP feature vectors established from the first experiment after subdividing the images into nxn blocks for n=1,2,…,9.

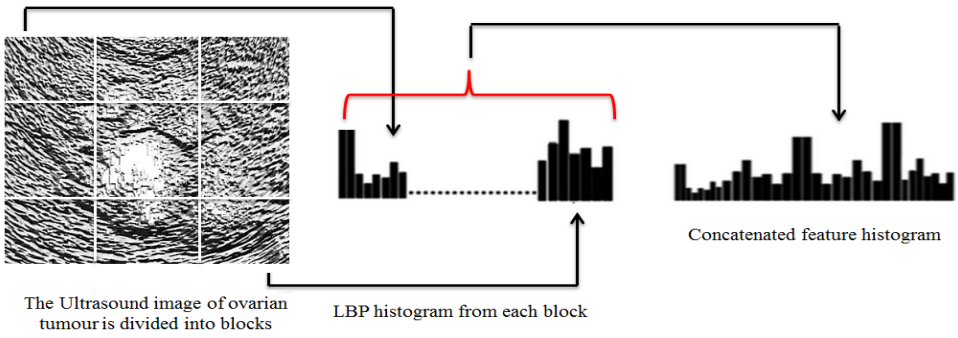


Figure 5. 9: Example of blocked based LBP histogram for ultrasound image of ovarian tumour

Figure 5.10 compares the classification results together with sensitivity and specificity rates based on LBP 256, 59, 58 bins extracted from each block of a 4x4 image subdivision. The result demonstrates that the simple LBP 256-bin histogram scheme outperforms the other two in terms of distinguishing between benign and malignant ovarian tumour. This confirms that the frequency of each of the 256 bins is important for discriminating these two classes of tumours. Accordingly, grouping all non-uniform bins under one bin seems to have some negative impact on the overall accuracy. This also explains the observation that the LBP-58 scheme slightly outperforms LBP-59, in terms of diagnosis accuracy.

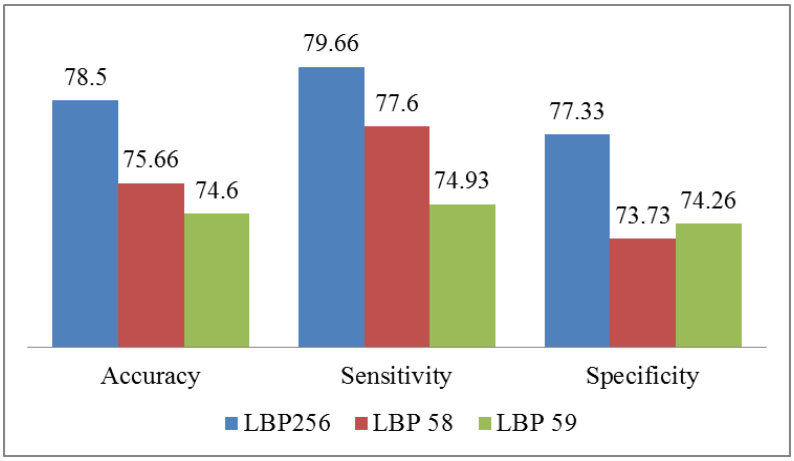


Figure 5. 10: Accuracy rates of classification using LBP 256, 59, 58 bins

The results of the second experiments conducted to test the performance of various localised versions of the simple LBP feature vectors, for different image subdivision methods, are shown in Figure 5.11 These results confirm that capturing the local LBP 256 bins features yield better diagnosis accuracy compared with the global version (i.e. 1x1 subdivision scheme). The figure reports that the overall diagnosis accuracy was improved from 72.6% using the whole image to 78.5% using 4x4 blocks. The figure also illustrated that more blocks lead to decreasing the overall classification accuracy.

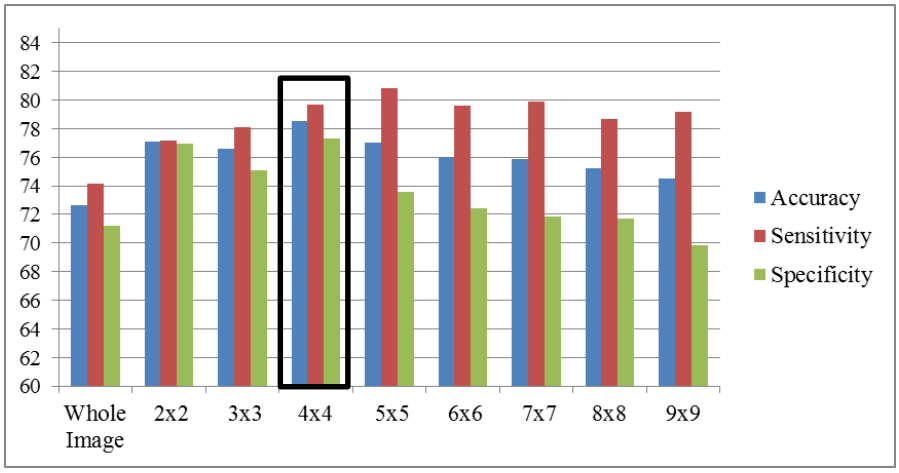
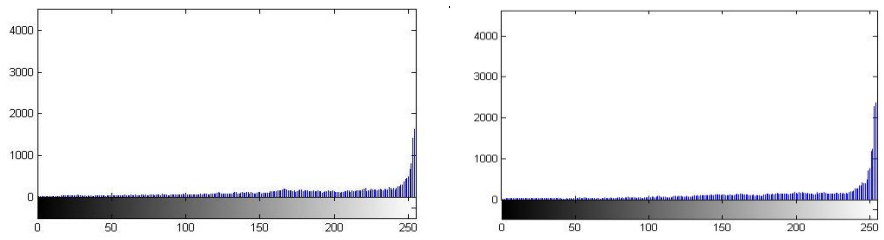
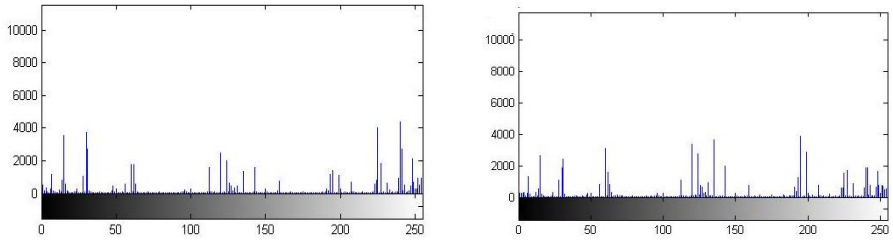


Figure 5. 11: Effect of blocking on the LBP image based on 256 bins

A close analysis of the results, presented in this section and previous sections, reveals that some images that were not successfully classified using the image histogram feature have been correctly classified using LBP and vice versa. For example, Figure 5.12 presents of the histogram of colour intensities and that of LBP 256-bin of two images of different tumour classes, one benign and another malignant. Clearly, intensities histograms are similar but LBP histograms are different. This is an important observation and suggests that features fusion could improve diagnosis accuracy. This will be investigated further in Chapter 6.



(a) (b)



(c) (d)

Figure 5. 12: Intensities histograms Vs LBP Histograms for different US images of ovarian tumour (a) Histogram of benign tumour (b) Histogram of malignant tumour (c) LBP Histogram of benign tumour (d) LBP Histogram of malignant tumour

5.2.2 Novel Feature Extraction Method based on Fast Fourier Geometrical Features FFGF

The results in the previous sections, on the automatic identification of ovarian tumours, have demonstrated beyond any doubts that the various feature vectors extracted from the spatial domain of the US scan images can significantly contribute, to better diagnosis when used as an aid to the medical experts. The lowest accuracy rate reported by the variety of schemes exceeds 65% which is certainly way above the capability of the random guess. These schemes mostly extract local texture data from the spatial domain. The fact that image frequency domain provides a rich and tested source of texture information was the main incentive to extract frequency domain feature vectors, we have developed the novel and elegant FFGF feature vector extracted from the binarized spectrum of the FFT of US images.

The FFT spectrum by itself only provides information about the directions of the dominant discontinuities in the image (e. edges and other geometric features) without giving any spatial information about the position of such discontinuities. The discontinuities are indicated by the highlighted rays that seem to be radiating from the central frequency at (0,0) which represents the total *energy* (also known as the DC component in the image). Binarizing the FFT septum image using some sensible thresholds produces an elliptic shape like object in the centre whose axes and size are good indicators of the orientation of dominating textures in the original image. Accordingly, we propose the following Fast Fourier- based Geometric Features (FFGF) to model the shape of low frequencies spectrum extracted by the following steps:

Step 1: Pre-process the US scan image using the method explained in (Section 4.3).

Step 2: Transforming an image into FFT frequency domain and compute its spectrum.

Step 3: Binarise the FFT power spectrum, using a sensible threshold.

Step 4: Determine the best ellipse fit of the highlighted shape in the centre of the binary spectrum image.

The regionprops function in Matlab is used to fit an ellipse to the spectrum by matching the normalised second central moments as a threshold (Haralock and Shapiro 1991). Then from the fitted ellipse, we extracted a set of three features FFGF feature vector:

FFGF() = [*Major axis* *Minor axis* *Ellipse Area*].

Figure 5.13 illustrates the block diagram of our FFT geometric feature extraction method.

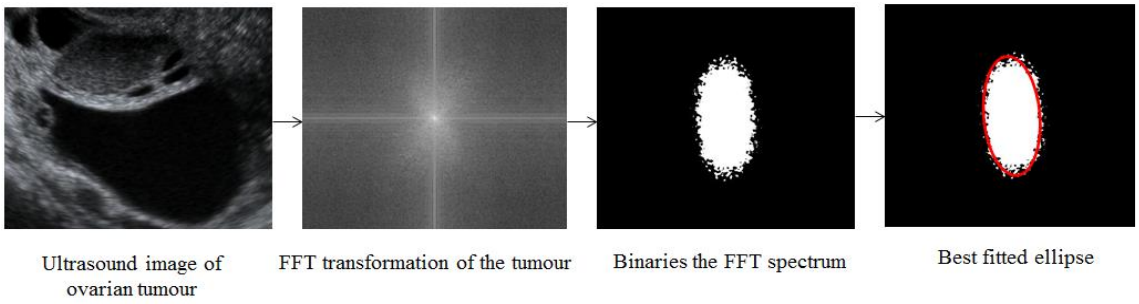


Figure 5. 13: Block diagram of the proposed method based in frequency domain

We conducted experiments using the same evaluation protocol, used above, and we built the linear SVM classifier to obtain the support vectors for the FFGF features extracted from the same training set of US images of ovarian masses. Figure 5.14 displays the result of automatic identification of the probe set of images in terms of accuracy rate and the corresponding sensitivity and specificity measures. These results demonstrate the significant benefits of extracting feature vectors from the FFT domain in discriminating benign and malignant tumours. The significance of these results stems from the fact that the extracted frequency domain features are even less accessible to medical experts through visual inspection than the other three sets of spatial domain feature vectors. More importantly, that it takes only three basic frequency domain features to achieve the accuracy of above 85%. This would certainly provide a valuable contribution to medical tools with very little added cost. The figure also shows a significant performance in terms of Sensitivity and Specificity measures.

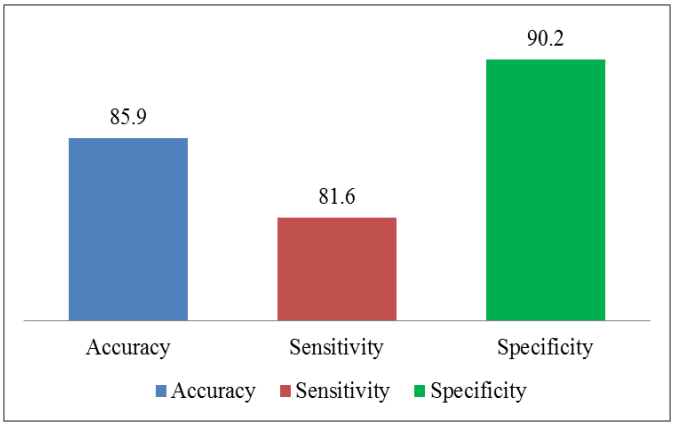
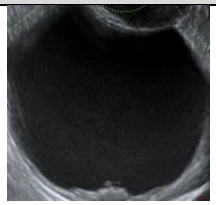
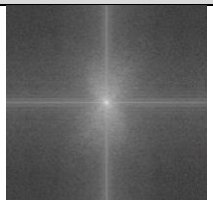
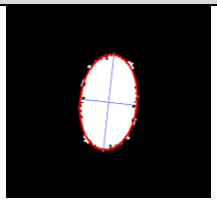
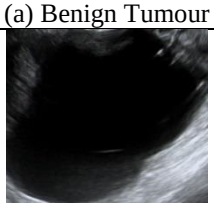
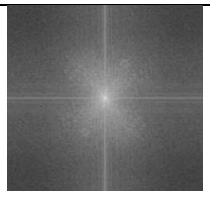
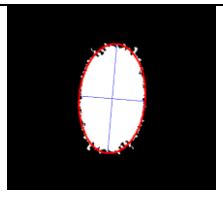
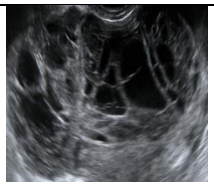
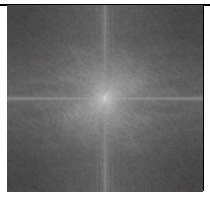
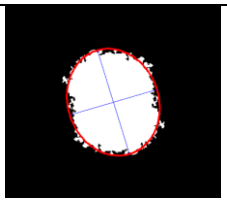
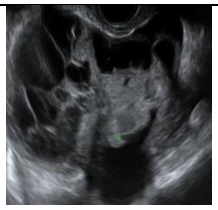


Figure 5. 14: Classification result based on FFGF using SVM classifier

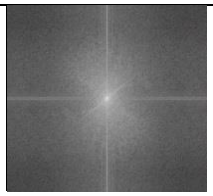
The rationale behind the proposal is that all detailed information including the texture of the image is represented by low frequencies around the centre of the spectrum. It is medically known that US scan images of malignant tumours tend to contain more details and complex structures compared to the much simpler images of benign tumours. Images listed in Table 5.1 confirmed this observation when we modelled the shape of low frequencies spectrum by the FFGF features. The table shows that malignant images tend to have bigger/fatter ellipses i.e. more details and more complex structures compared with those of benign cases.

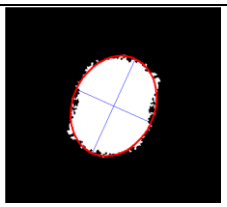
Table 5. 1: Comparing Spectrum of US scans of ovarian benign and malignant tumours

Enhanced ROI	FFT Spectrum	Binary spectrum with best fitted ellipse
 (a) Benign Tumour		
 (b) Benign Tumour		
 (c) Malignant		



(d) Malignant





To appreciate the implications of these results we present in Figure 5.15 a comparison of the performance of the FFGF based method with that of the other spatial domain features discussed in the last few sections. The figure clearly shows the proposed FFGF features were able to capture the differences between the two classes of tumours, particularly in the success in correctly identifying benign tumours.

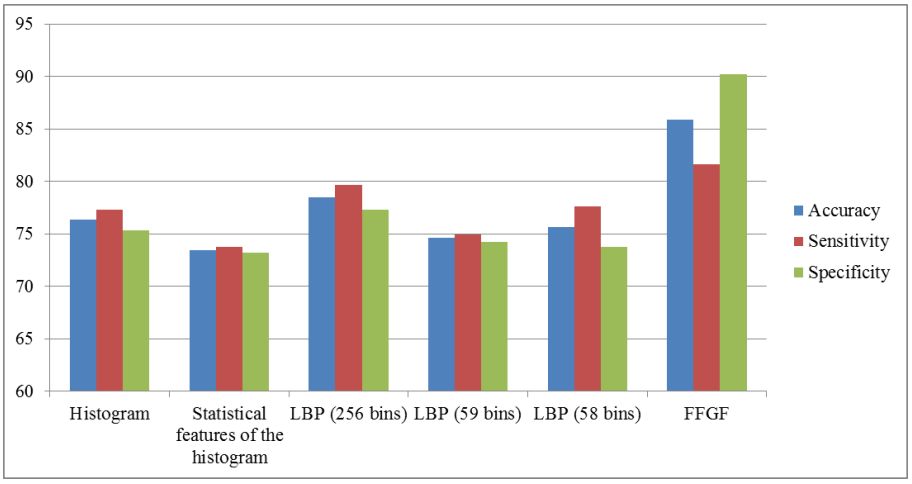


Figure 5. 15: Classification results using all feature vectors (SVM)

These results also indicate that fusion may provide even more improved accuracy. This will be elaborated on in Chapter 6. It would be interesting to see how effective this newly proposed feature can be in our other area of interest, i.e. identification of miscarriage.

5.3 The Proposed Method for Miscarriage Identification from US Scans of the Gestational Sac

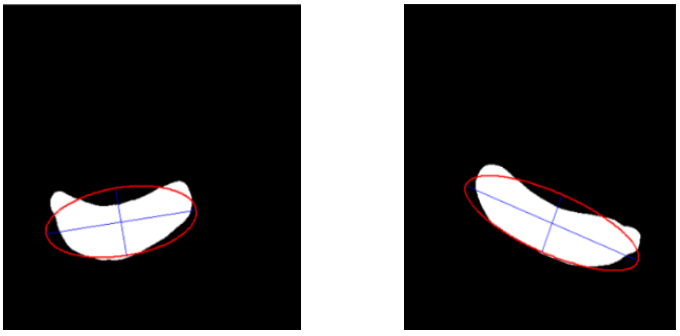
In this section, we investigate automatic schemes for identification of miscarriage from US scan image of the GS in pregnant women. Gynaecologists manually extract certain geometric parameters of the GS and examine various aspects of the GS such as the presence/absence of the Yolk sac and embryo heartbeat before deciding whether a miscarriage has occurred or not. The contributions of this section are two folds. We first attempt to automate the extraction of the GS geometric parameters currently obtained manually by gynaecologists,

plus a few other GS geometric parameters that have not been used and develop computer-based schemes for classification of miscarriage from US scan of the GS. We also test the possibility of extracting the various image-texture based feature vectors, introduce in section 5.1.5, from US scan of the GS for identification of miscarriage.

Again we point out that we first enhance and segment the gestational sac using the procedures described in Chapter 4, Section 4.2, we prepared the segmented sac to the features extraction step. Once the various existing and new features are extracted, we conduct performance testing experiments, with the US scan of GS dataset described in (Chapter 3, Section 3.7) but this time using the kNN classifier with the Euclidean distance. The effect of using different classifier will be discussed in Chapter 6.

5.3.1 Geometrical Measurement Features from Gestational Sac

As explained earlier, each GS is viewed in two perpendicular planes. The GS is usually round in shape at the early pregnancy, but as the sac grows it becomes more elliptical. Therefore, our algorithm aims to find the best-fitted ellipse for the segmented GS for each plane. The regionprops function in Matlab is used to fit an ellipse to the sac region (Haralock and Shapiro 1991). This function returns four parameters: Major and Minor axes, Centroid and Orientation which is the angle between major and minor axes, as shown in Fig 5.16 Assuming that the GS has an ellipsoidal shape in 3D, the three principal axes of the ellipsoid has traditionally been estimated by the major axis (A), minor axis (B) of the ellipse from the sagittal plane and the major axis (C) from the transverse plane. After that, we define the following geometric set of features from each GS:



(a) (b)

Figure 5. 16: Automatic best fitted ellipse on segmented GS with the Major and Minor axes (a) Segmented GS in sagittal plane (b) Segmented GS in transfer plane

1. **Mean Sac Diameter (MSD):**

This is simply the average of three principal diameters along the A, B, and C axes, from the two planes of the GS.

2. **Volume:**

The GS volume can be estimated using the three principal diameters (Figure 5.17) as follows:

$$V = \frac{1}{8}.\frac{4}{3} \cdot \pi \cdot A \cdot B \cdot C \quad (5.12)$$

we take the cubic root of the volume as a volume measure.

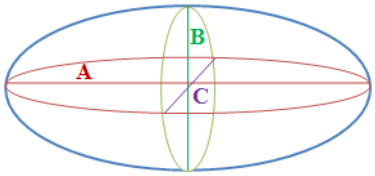


Figure 5. 17: An ellipsoid shape with its three principal axes

3. **Perimeter:**

This is calculated by counting the number of pixels around the boundary of the GS, and then taking the average of the perimeter from both planes to produce a single perimeter measure.

Besides the above features, in some cases, gynaecologists estimate his/her decision based on the shape of the GS (i.e. irregular border of the GS). Here, we define such features as image parameters, and test their performance separately to see if any benefits can gain. These features are:

4. **Circularity:**

A very common shape factor is the circularity, if circularity 1 indicates for circle shape. This feature depends on the area and perimeter of the object. We calculated the average of the circularity from both planes, to produce a single circularity measure.

$$\text{Circularity} = 4\pi * \frac{\text{Area}}{\text{Perimeter\^2}} \quad (5.11)$$

5. **Compactness:**

This factor shows the roughness of the GS boundary. The smaller value indicates the most roughness. Averages of both planes are taken to produce one compactness measure.

$$\text{Compactness} = \frac{\text{Peremtier}^2}{\text{Area}} \quad (5.12)$$

6. Solidity:

Show irregularity of object border, the value of solidity will be small if the border is irregular. From the average of both planes we produced one solidity measure.

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Area}} \tag{5.13}$$

7. Eccentricity:

The eccentricity is the ratio of the distance between the foci of the ellipse and it is major axis length the value is between 0 and 1. We calculated the average of the eccentricity from both sections to produce a single eccentricity measure.

8. Maximum Diameter:

After we found major and minor for each section, then chose the maximum diameter between four measurements.

The MSD is the one measure that is common between the above automatically extracted features and the ones used by gynaecologists. Therefore, it is natural to compare between Automatic MSD and the Manually measured MSD. This also helps illustrate the effectiveness of our automatic segmentation. Figure 5.18 presents a scatter plot of all 184 images as points along the automatically computer estimated and the manually measured MSD dimensions. The figure shows that the values of automatic MSDs are closely correlated to the manual MSD (R-square = 0.98). Also, the approximation line across the points has an angle close to 45% with points located near to the line on both sides, indicating that (a) the automatic measurement, in general is close the manual measurement, and (b) there is no systematic bias in the automatic measurements. We should also emphasise on the point that there lacks of a "ground truth" in manual measurement that we can rely on to correctly identify the margins of error made by the automatic measurement.

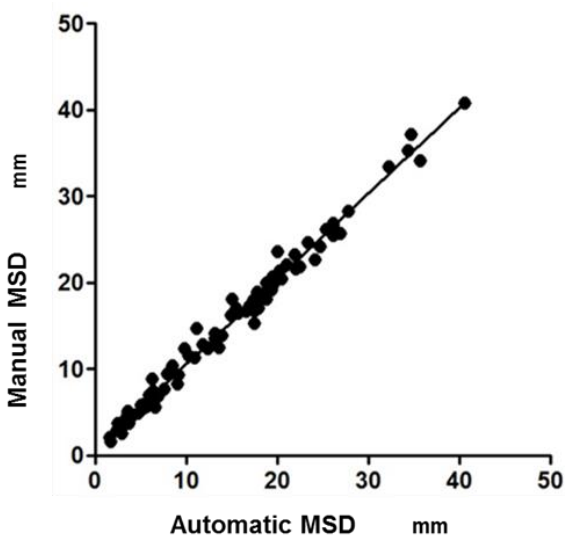
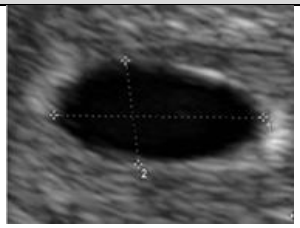
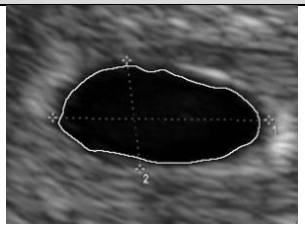
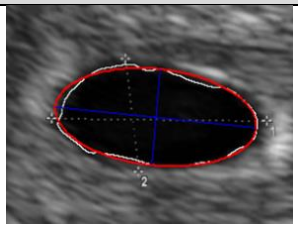
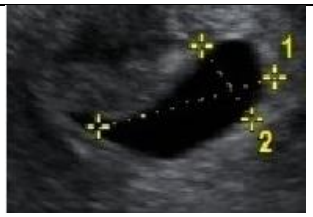
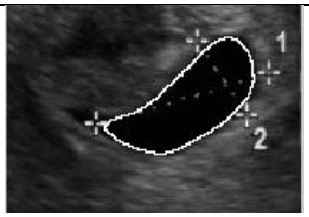
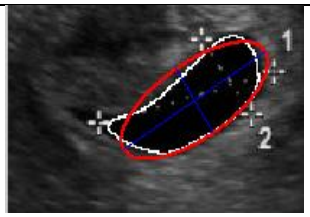


Figure 5. 18: Differences between manual and automatic measurements of MSD (R- square= 0.98)

Table 5.2 presents the images of the two example GSs with manual landmarks as well as automatic measurements of the two diameters in the sagittal plane (d1 and d2). The Figure shows that both types of measurement lead to margins of error from the actual size of the GS. The automatic method always attempts to search for the maxima when estimating the diameters of the best-fit ellipsoid, whereas human operators do not and can be very subjective.

Table 5. 2: Examples of manual vs. automatic measurements

Manually measured d1 & d2	Detected Border of GS	Estimated d1, d2 vs Manual d1, d2
		
(a) A regular shaped GS where manual and automatic measurements are close		
		
(b) An irregular shaped GS where manual and automatic measurements are different		

5.3.2 Performance Testing Experiments

For each US image, the above-listed features can be treated as separate parameters or as an attribute of a feature vector for diagnosis purposes. The aim of this section is to test the performance of automatic identification of miscarriage from US images of the GS in pregnant women. The experimental dataset for this task arrived in two batches (94 in the first batch and 90 images in the second). The second set differs from the first set due to the difference in source scan machine, time of the scan, and the observers. Accordingly we conducted two sets of experiments. The main purpose behind the first set of experiments is to train the proposed methods on one dataset and test it on another independent dataset whereas the second set of experiments was conducted on a combined dataset using the cross-validation protocol.

1. **First Experiments:**

This experiment was designed to train our automatic identification schemes by using the first set (94 images) and testing the accuracy using the new set (90 images) as test images. In this experiment, the old set consisted of 79 PUV cases and 15 miscarriage cases, while the new set consisted of 78 PUV cases and 12 miscarriage cases. We used separately each of the first three geometrical features defined above extracted from the best-fitted ellipse imposed on the gestational sac area and then we concatenated the three features to see the effect of combined features on classification results.

Figure 5.19 shows the result of this set of experiments. Using the automatic MSD feature alone, an overall accuracy near to 99% with sensitivity (miscarriage) of 100% and specificity (PUV) of 98% was obtained. This is comparable to the accuracy of 98.8 % (sensitivity 100%, and specificity 97.6%) when used the volume feature alone, and an overall accuracy of 97% (sensitivity of 100% and specificity of 94.1%) when used the perimeter feature alone. Combining all three features does not seem to lead to better classification results. This is because the three types of features are highly correlated.

It should be said that the perimeter remains as a robust feature which is automatically measurable from each single image of the GS and still gives the relatively good performance, and hence should not be easily dismissed. Moreover, its roughness can also be used as added information in unsure cases.

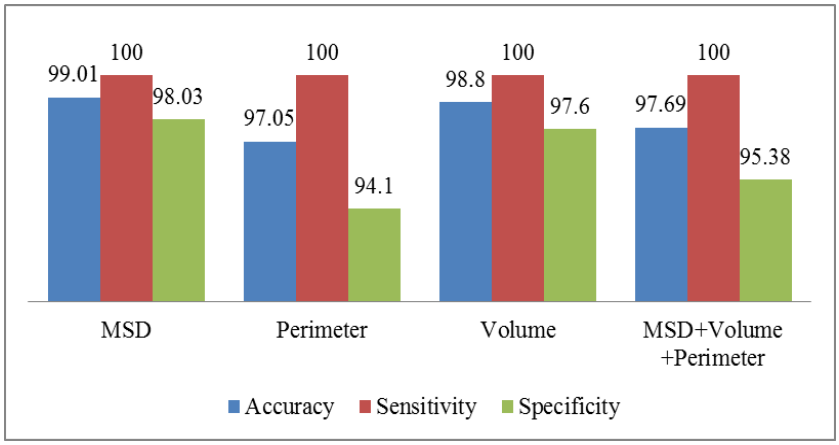


Figure 5. 19: Experiment 1: Miscarriage classification using MSD, perimeter, volume and combine all three

2. Second Experiment (Combining Dataset 1 and Dataset 2):

We conducted another set of experiments by combining the two datasets 1 and 2, i.e. we have 184 ultrasound image of GS were used as a new set (157 PUV cases and 27 miscarriage cases). The testing protocols are all described in Chapter 3, Section 3.6.1.2. From Figure 5.20 is appear that the accuracy result is marginally higher when using MSD (99.2%) and Volume (99%) while for the Perimeter we achieve a marginally lower accuracy of 96.6%. Combining all three features yields accuracy of 98.8%, which is an improvement of more than 1% on the results of experiment 1, but it is still less than MSD.

In short, the results from the second set of experiments on the effectiveness of the extracted features give similar readings as those from the first experiments confirming that the accuracy rates are not influenced by the differences between the two data sets, i.e. the system works well regardless of the scanning machine, the time of the scan and operator skill. While the sensitivity remains high for MSD and volume features, the specificity for the perimeter feature is reduced marginally. At the same time, the sensitivity for each type of feature has improved to 100% while sensitivity dropped slightly.

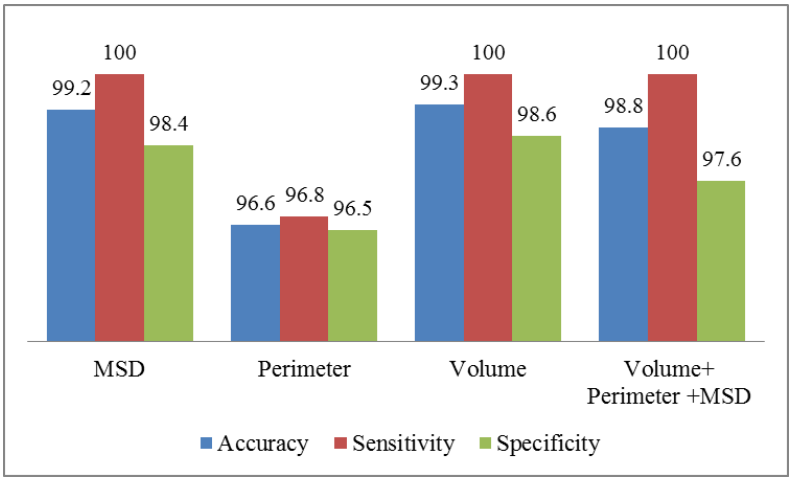


Figure 5. 20: Experiments 2: Miscarriage classification using MSD, perimeter, volume and combined all three features

We now investigate the performance of the additional six geometrical features defined above (i.e. area, compactness, circularity, solidity, maximum diameter and eccentricity) for identification of miscarriage cases. The results of the classification tests, using the same setup as in experiments 1 and experiment 2 above, on these features are shown in Figure 5.21 to be significantly inferior to the performance obtained by the other three features. These experiments confirm that these features have little additional diagnostic value for miscarriage cases. The rather high performance of the Area, and to some extent the Max Diameter, features reflect the fact that these parameters are closely related to the various diameters. The reason for the poor performance of the other features may well be due to the images in our data set where most sacs, miscarriage or PUV have regular and smooth borders. However, more images of various border characteristics are needed in future investigations to draw solid conclusions

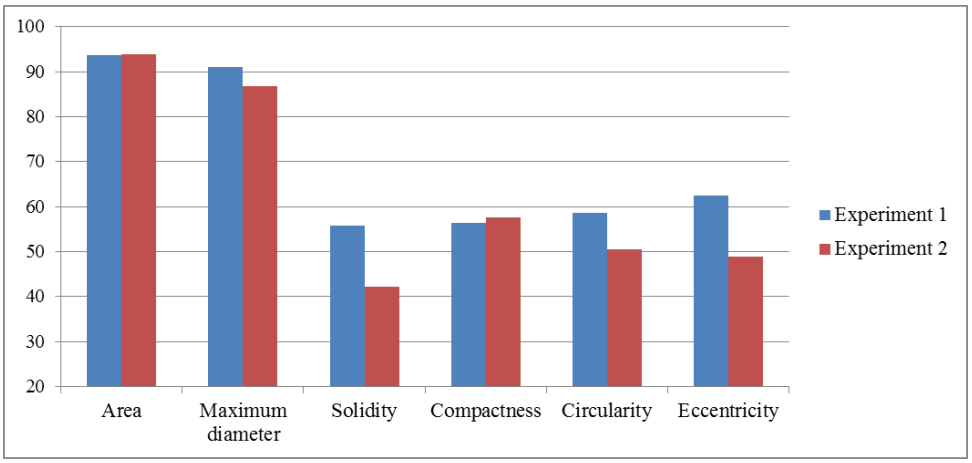


Figure 5. 21: Performance of the six additional features, using kNN classifier

Close examination of the misclassified PUV images, i.e. false positives, reveals that their MSDs varied from 20 to 25 mm which is very close to the borderline between PUV and miscarriage cases. Our expert collaborator, who prepared the dataset, confirmed that all these borderline PUV cases eventually became cases of miscarriage in follow-up diagnoses. Table 5.3 shows the borderline examples that were diagnosed by the expert compared to the automatic method, confirming that our automatic method enabled more accurate and advance prediction than the manual method in this case. More experiments are needed to consolidate this finding.

Table 5. 3: Represents the Manual vs. Automatic diagnosis

Image ID	Manual Diagnosis			Computer Based Diagnosis	
	MSD	First Diagnosis	Ultimate Diagnosis	MSD	Automatic Diagnosis
17	19.15	PUV	Miscarriage	19.18	Miscarriage
67	22.67	PUV	Miscarriage	24.06	Miscarriage
46	15.30	PUV	Miscarriage	17.39	Miscarriage

5.3.3 Image Textural based Feature Vectors for Identifying Miscarriage Cases

Having demonstrated that the first three automated geometric features of the GS have almost perfect performance, we decided to investigate whether the image based features extracted from the spatial/frequency domain of the US scan images of ovarian tumour tumours can add any benefits for the identification of miscarriages.

We extracted the FFGF, the LBP and the Grayscale Histogram features as texture feature space of the US scan images of the GS from the two sets of images used in experiment 1 and experiment 2, above. The experimental results demonstrate a rather modest discriminating power of these image-based features. The results shown in Figure 5.22, are very poor compared with the performance of the three geometrical features tested in the last section.

These results confirm that miscarriage identification is entirely dependent on the shape and size of the GS sac. However, the texture inside the GS tends to be very similar in both cases (PUV and miscarriage).

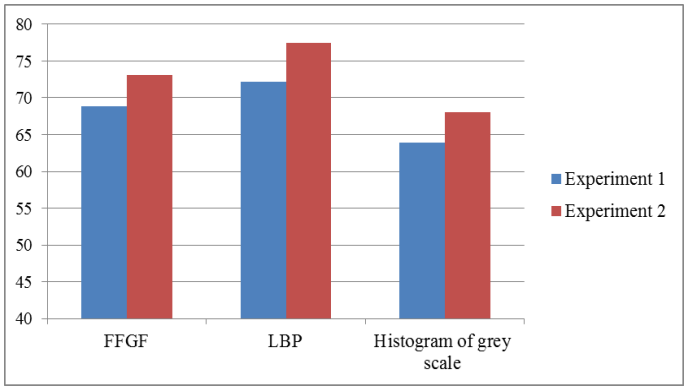


Figure 5. 22: Accuracy of identification of miscarriage and PUV cases for all three feature vectors

5.4 Summary and Conclusions

This chapter was primarily devoted to select and extract features from pre-processed US scan images of gynaecological that have a high capability in discriminating different types of abnormalities. In our research framework, this is an essential component and identifying appropriate features is an application-dependent effort. Hence, we dealt with the identification of miscarriages separately due to the fact that medical experts use the well-defined set of features.

We have successfully developed and demonstrated the reliability of automated procedures to extract several features that represent various GS dimensional measurements, few of which are manually determined by a radiologist and relied upon by gynaecologists to identify miscarriage cases. The experimental work conducted to test the performance of automatic diagnostic schemes based on these features have shown beyond any doubt excellent accuracy rates when the kNN classifier used, with the MSD achieving in the high 90's. The choice of kNN is based on the widely accepted knowledge of its suitability for low dimensional feature spaces.

For the other gynaecological application of diagnosing Ovarian Tumours, our investigations where more motivated by the well-known practices in image based pattern recognition applications and was guided by medical expert advice on the importance of texture in distinguishing between benign and malignant tumours. Image analysis theory provides a range of tools to determine image texture information both in the spatial and frequency domains. We adopted a number of feature vectors well-known for their ability to encapsulate local as well as global image texture information. In the spatial domain, such features include image/sub-image histograms, statistical moments of grey scale, and local binary patterns (LBP). Relevant frequency domain features can be defined using different frequency analysis

transforms, but the FFT spectrum of the image is known to provide a rich source of global image texture information. We developed an innovative new condensed low dimensional feature vector obtained from a special threshold-based binarised FFT spectrum. This feature encapsulates the image texture information that depends on all image pixel values. We denoted this three-dimensional feature vector, extracted from an ellipse shaped object, by FFGF. The experimental work conducted to test the performance of automatic Ovarian tumour diagnostic schemes, based on these various spatial/frequency domain image-texture features using the SVM classifier, have demonstrated that each one of these schemes has the satisfactory-to-considerable power of discriminating Benign from Malignant tumours that cannot be attributed to chance. Interestingly, our innovative three-dimensional FFGF feature vector has significantly outperformed the spatial domain features, achieving more than 85% accuracy. Again the choice of SVM, in this case, was based on the widely accepted knowledge of suitability for high dimensional feature spaces. Note that, the FFGF encapsulates global image-texture that is extracted from the high dimensional FFT spectrum.

The successful outcome from this chapter investigation indicates the potentials of using machine learning to provide support for vital health decision-making in the highly specialised branch of biomedical image analysis for gynaecological abnormalities. This will be the main focus of the investigations in the next chapter.

CHAPTER 6

FUSION BASED CLASSIFICATION OF GYNAECOLOGICAL US IMAGES SUPPORTED BY LEVEL OF CONFIDENCE

In the previous chapter, we investigated the automatic extraction of a mix of different existing and new sets of features from US scan images to characterise ovarian tumours and gestational sacs. We illustrated the performance results of these sets of feature vectors by using one specific classifier (SVM for ovarian tumour classification and kNN for miscarriage identification). Based on the research rationales outlined in Chapter 3, the next step of our research is to investigate the effectiveness of using multiple classifiers and fusion scheme, for improved diagnoses of ovarian tumour states and miscarriage cases. The fusion investigations will be confined in fusing at different levels, the outcome of the SVM and kNN classifiers only for each of the extracted single/concatenated feature vectors. Moreover, in this chapter we introduce the concept of level of confidence to the diagnosis outcome. The specific contributions in this chapter include:

- Improving the performance of the automatic binary classification based on various fusion approaches in multi-feature schemes.
- Introducing the concept of level of confidence in the classification decision relevant to the two applications, namely miscarriage diagnosis and ovarian tumour identification.
- Fusing different sets of features at the decision level supported by the level of confidence measures, and introducing a Not Sure decision to flag the difficult cases that could have been misclassified.

In Section 6.1, we review the literature about classifying different diseases using ultrasound images. In Section 6.2, we extend the work in Chapter 5 and test the performance of the various sets of feature vectors for both gynaecology abnormalities when using the other classifier (i.e. kNN for ovarian tumour classification and the SVM classifier for the identification of miscarriage). Automatic decisions made by these two classifiers rely on meeting some criteria that depended on a distance function or a similarity score, and the distribution of the computed distances/scores could provide some indication of the reliability/confidence of/in the classification. In Section 6.3, we introduce the concept of confidence level and models for quantifying confidence in decisions made by kNN and SVM classifiers. In section 6.4, we test the performance at three levels of confidence of (1) the

kNN classification for the various GS miscarriage application schemes, and (2) the SVM classification for the various Ovarian Tumour application schemes. In Section 6.5, we investigate various types of multi-schemes fusion to improve the accuracy of SVM classification of Ovarian Tumour US images. This investigation includes fusion at the feature level, at score level and decision level. The decision level fusion includes a novel approach which would be based on levels of confidence, and demonstrate the effectiveness of this approach with experimental results for US images of Ovarian Tumours. Section 6.6 summarises the chapter with concluding remarks.

6.1 Classification of Ultrasound Images - Literature Review

In the medical domain, different classification techniques have been applied for different medical diagnosis purposes. In recent years, there have been increasingly more research works in the area of ultrasound image classification for different diagnostic purposes. As we see next several research projects, have been reported in the literature on the classification of ovarian abnormalities, but to our knowledge, no existing work has been reported in the literature the area of automatic GS classification.

A multi-class SVM classifier was applied on histogram moments with Grey Level Co-Occurrence Matrix (GLCM) as features to classify three different types of ovarian cysts: Simple Cyst, Endometrioma, and Teratoma (Sohail, Rahman et al. 2010). The method was tested on 478 ultrasound images of ovarian cysts with a claimed average accuracy of 86.90%. More work on ovarian tumour classification was done using the Decision Tree classifier on Higher Order Spectra (HOS) textural features (Alcázar and Jurado 2011). This technique was validated using 1000 benign and 1000 malignant images obtained from 10 patients with benign tumours and 10 with malignant tumours respectively. With a 10-fold stratified cross-validation, a high accuracy of 95.1% (sensitivity of 92.5% and specificity of 97.7%) was reported. The limitation of this work is that the images are obtained from just 20 patients and with limited pathology. Later, the same team of researchers proposed a new solution in (Acharya, Mookiah et al. 2013) based on Hu's invariant moments, Gabor transform parameters and entropies. These features were first extracted from 2D colour Doppler and 3D ultrasound images of ovarian tumour. Then probabilistic neural network (PNN) classifier was then applied to discriminate between benign and malignant tumours. The proposed solution was evaluated using a 10-fold cross-validation on 1300 benign images and 1300 malignant images obtained from 10 patients with a benign tumour and 10 with a malignant tumour, and

an average classification accuracy of 99.8% (sensitivity of 99.2% and specificity of 99.6%) was reported. Again this work was done based on the abundant supply of images but from a very limited number of patients (20).

In a research project about breast tumour classification in (Liao, Wan et al. 2011), three types of classifier SVM, ANN and kNN were evaluated on 321 ultrasound images to differentiate benign from malignant breast lesions based on the moment features that describe the contour of the breast lesions in sonography. The authors concluded that the performances of all three classifiers achieved the accuracy rate of over 80%, with SVM achieving the best accuracy of 86.9%. In a study on prostate cancer lesion classification (Moradi, Mousavi et al. 2009), SVM was used to classify the ROI's of 35 ultrasound images, and 12 feature components representing texture were extracted. Four of the features were the statistical moments of the pixel intensities (i.e. mean, standard, skewness, and kurtosis) and the remaining eight features were from the co-occurrence matrices (namely correlation, energy, contrast, and homogeneity in the 0° and 90° directions). An overall accuracy of 91.7% (sensitivity of 86.6%, and specificity, of 94.7%) was reported.

In the next section, we expand our work reported in Chapter 5, using two different classifiers (SVM and kNN) for the automatic classification of the ovarian tumour as well as identification of miscarriage from US scans of the GS. We shall do that using the variety of feature vectors. In our work, the numbers of scan images are smaller than those reported above, but the numbers of patients are considerably more, in ovarian tumour dataset we have 187 images which these belong to 177 patients, whereas in GS dataset we have 184 patients.

6.2 Performance of Different Classifiers for Gynaecological US Images

6.2.1 Classification of Ultrasound Images of Ovarian Tumours

We followed the experimental protocols as explained in Chapter 3, Section 3.7 to evaluate the performances of our two chosen classifiers, i.e. kNN and SVM using the different individual and multiple features as described in Chapter 5. Except for the FFGF feature scheme, for all other schemes, corresponding features extracted from a 4x4 block subdivision of the images are concatenated.

In this subsection, we look at the classification of ultrasound images of ovarian tumours first. For simplicity we set k=1 for the kNN classifier, Figure 6.1 shows the overall classification accuracy rates achieved by each single feature scheme with both SVM and kNN classifiers. The figure clearly shows that the SVM consistently outperforms the kNN classifier on all

types of features for classifying benign and malignant tumours. The performance differences between the two classifiers range from 8.6 % for the statistical histogram features to 16.7% for the FFGF features. Since most types of feature vectors are of relatively high dimensions, the good performance of SVM in high dimensional feature spaces partially explains its advantage over the kNN classifier. However, it should be noted that SVM also has the best performance over kNN for the FFGF features that are of significantly low dimensionality (three attributes only).

In other words, the rather very low three-dimensional image-texture based features extracted from the Fourier transform spectrum are better separated by the linear hyperplane of the SVM than the Euclidean distance of the nearest neighbour. This raises some questions about and exceptions to the validity of the widely accepted characteristic of the SVM as being more suitable for high dimensional feature space. A possible justification of the exception in this case relates to the way the FFGF features encapsulate the information in their source. The FFGF is extracted from the FFT spectrum image each pixel value of which represents the texture information throughout the *US image* that proportionally highlighted by the pair of waveform frequencies indicated by the spectrum pixel position (u,v). This makes FFGF as a compressed version of the entire image texture with little loss of information, i.e. FFGF acts as a very effective dimension reduction procedure.

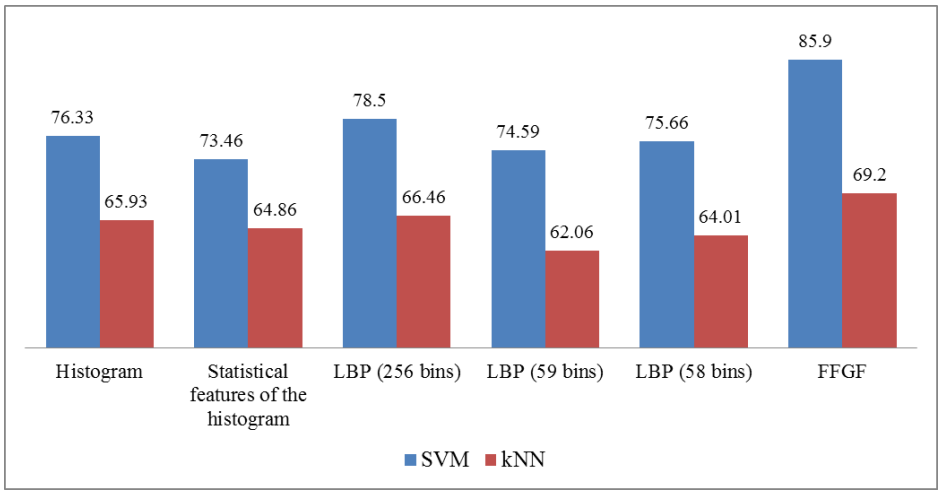
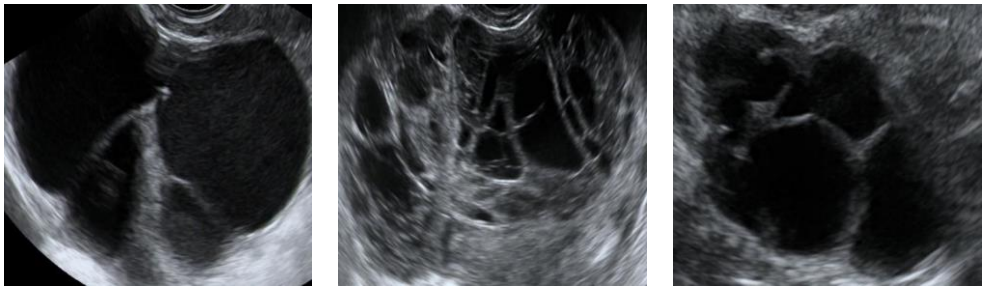


Figure 6. 1: Ovarian tumour classification accuracy by SVM and kNN

The classification results still demonstrate the reasonable success of single feature schemes but still not optimal. The highest level of overall accuracy is only 85.9% for the FFGF features, and the lowest level of accuracy is only 73.46% for the statistical features of the histogram. Either classifier could result in misclassifications. After closely examining the types of ultrasound images classified as well as sensitivity and specificity by each classifier,

we realise that both the SVM and the kNN tend to make correct predictions for benign tumours of relatively simple textures. However, for malignant tumours of complex textures and structures, the SVM makes better classification decisions than the kNN classifier, but both the SVM and the KNN also makes errors with benign tumours of lesser complex textures and structures. Figure 6.2 presents some example images together with the classification results by the kNN and SVM classifies on the extracted FFGF features. This interpretation tends to suggest that fusion of different classifiers with the same or different feature vectors may complement each other and make better diagnostic classifications (see further discussions on this issue later in this section). Testing the fusion of the two classifiers but with different feature vectors involve a great deal of experimental tests which will certainly be of serious interest especially for discriminating the various subclasses of malignant and benign tumours, which is outside the scope of the current project. In this chapter we shall focus on fusing one classifier with different feature vectors.



(a) (b) (c)

Figure 6. 2: Example images correctly classified or misclassified by SVM and kNN (a) Benign ovarian tumour correctly classified by both classifiers kNN and SVM (b) Malignant tumour correctly classified by SVM but misclassified by kNN (c) Benign tumour correctly classified by kNN but misclassified by SVM

6.2.2 Classification of Ultrasound Images of Gestational Sac

Here we expand the work reported in Chapter 5 to examine and compare the classification by the kNN and SVM classifiers on US scan images of gestational sacs for identification of miscarriage cases. We follow the same evaluation protocol as described in Section 3.7. As in Chapter 5, we conducted two sets of experiments. The overall accuracy rates are presented in Figure 6.3 and 6.4 for experiment 1 and experiment 2, respectively.

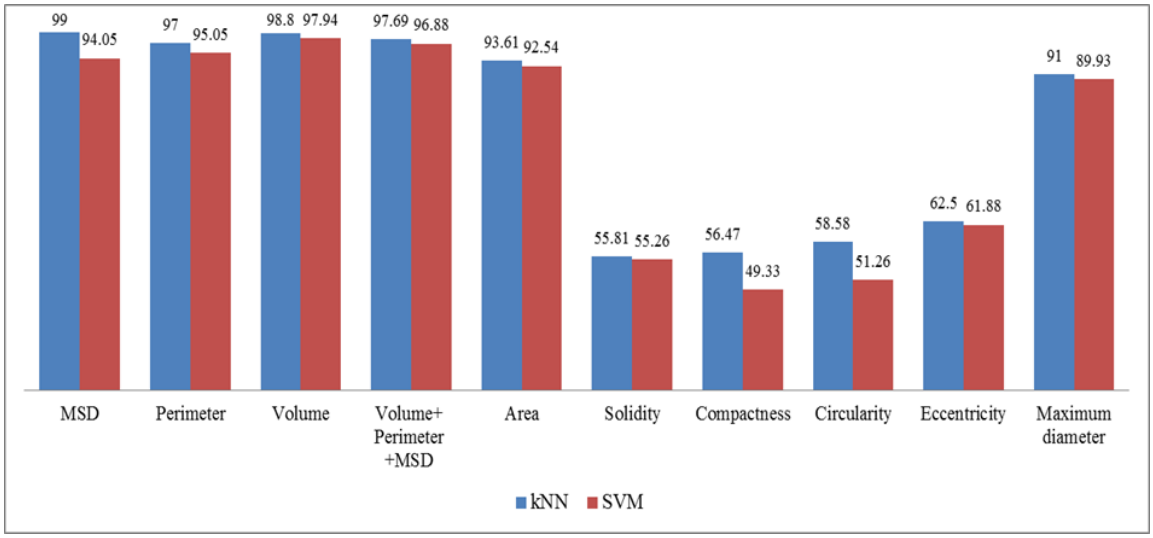


Figure 6. 3: Comparing performances of SVM and kNN for classifying GS - Experiment 1

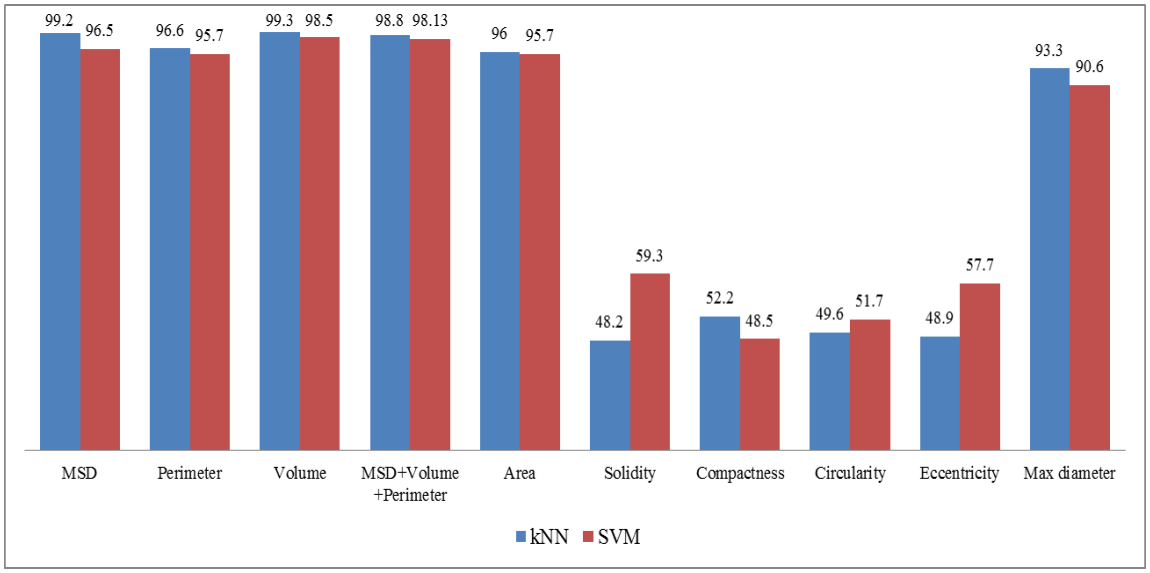


Figure 6. 4: Comparing performances of SVM and kNN for classifying GS - Experiment 2

One can easily observe that for this set of features the relative performance of the two classifiers is the opposite to what was observed for the classification of ovarian tumours. In fact, the kNN classifier gives consistently better though comparable results to those attained by the linear SVM for all types of extracted features in experiment 1, and for most of the types of features in experiment 2. However, for the image texture based features (Histogram, LBP and FFGF), the relative performance of the two classifiers are in the opposite direction. This is a further confirmation that SVM favours the high-dimensional feature vectors as well as the FFGF.

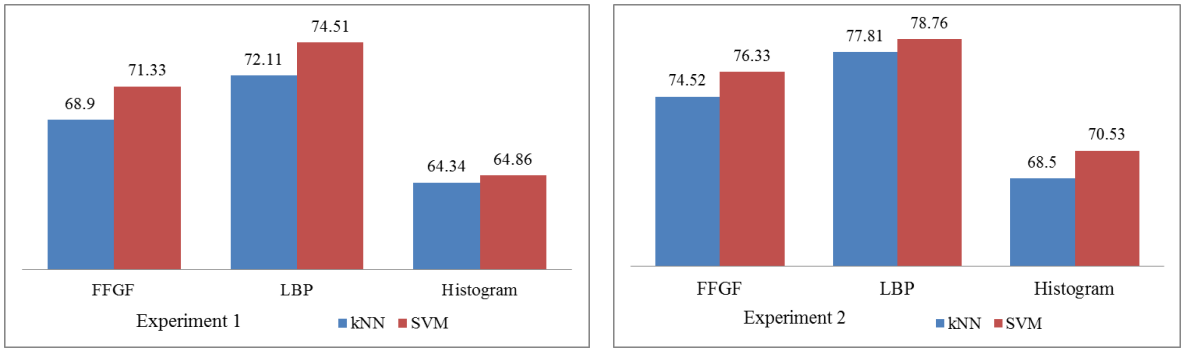


Figure 6. 5: SVM Vs. kNN for classifying miscarriage for image texture based features

However, the better performances of the SVM with the image texture based features are still way below the almost perfect performance of the kNN with the simple GS measurement features (i.e. MSD, Perimeter, and Volume). Note that such features are not applicable to ovarian tumours due to the fact that these tumours often consist of several sac-like compartments. The combined results can be summarised as follows:

- With very few exceptions, kNN marginally slightly outperforms the SVM in identifying miscarriages when features relating to the dimension of the GS are used, but this pattern is reversed when image texture based features are used.
- As established in chapter 5, the best performing feature vectors when kNN was used achieved the same performance ranking when SVM classifier is used.
- Comparing to the SVM classifier's relatively good performance for high dimensional feature vectors in the ovarian tumour diagnosis, the kNN classifier performs better than SVM for the low dimensional feature space in miscarriage case identification except for the FFGF feature vector.
- In the problem domain of miscarriage case identifications, the performance differences by different classifiers are much less than the performance differences by the features used. For most features, and in particular, the better performing ones, the accuracy rates of the two different classifiers are only marginally different.

Again these results provide motivation to consider fusion as a plausible approach to achieve better and more reliable solution of the miscarriage identification problem. However, this motivation is not strong as in the case of ovarian tumour classification since the performance of both classifiers are very similar and already very high.

6.3 Quantifying Confidence in Classification Decisions

In real life medicine, diagnosing any types of health related abnormalities using medical images (e.g. Ultrasound, MRI, X-ray, etc.), the observer first examines carefully the relevant scanned image and then makes the diagnosing decision using a well-established set of rules. However, different observers may come to different diagnostic decisions even when following the same procedure. In fact, diagnosis of medical conditions and the diagnosis accuracy very much depend on the experience of the observer. Very often, he/she will label the image as a negative or positive case, with the level of certainty. Sometimes, there can be cases where the observer alone cannot decide the diagnosis outcome when the level of certainty is lower than what makes the observer sufficiently sure in making a diagnosis decision. In such a scenario the observer often seeks opinions from other experts, to ensure raising the level of certainty regarding an outcome. In certain cases, however, combining different opinions may have a negative effect on the eventual diagnosis: instead of more certain, the observer becomes less certain. For instance, in some cases, two experts may reach opposite conclusions of the diagnosis; even each of them makes the decision with a high level of confidence. In such a situation, the diagnosis is considered inconclusive, and further medical checks such as surgery/biopsy on the patient are conducted. This informs us that medical diagnosis in practice is often not only concerned with whether a case is correctly diagnosed, but also how strong and reliable each diagnosis decision tends to be, and very often a final diagnosis decision making involves multiple decision makers. Therefore, it makes sense to investigate and design a quantitative confidence measure that could be embedded into the framework of automatic classification decisions that could also be exploited when fusing multi-classifiers for detecting the different classes of gynaecological abnormalities.

6.3.1 Level of Confidence in Classification

Generally in a space of data points of negative and positive classes, there exists a boundary that separates data points of one class from data points of another class. Classification models tend to locate or estimate such boundaries between training examples. At classification stage, the further away from the boundary a test data point is, the higher the level of confidence that could be associated with the decision about the class that the data point belongs to. On the contrary, the nearer to the boundary the data point is, the less confident the decision about the class to which the input data point belongs. In an extreme case when the data point is exactly

located on the boundary, the confidence is the least: it is most uncertain about which class the data point should belong to, i.e. a not sure case.

Decision strength has been embedded in many data mining solutions. In association rule mining, for example, confidence is defined as a posterior probability between the support of a rule and the support to the condition part of the rule (Agrawal R 1993). In Bayesian classifiers, the differences among the conditional probabilities of the classes that the data point belongs to can be taken as an indication how strong the final decision of taking the most probable class is (Pang-Ning Tan 2005). In decision trees, the proportion of the majority class examples at a leaf node can also show the strength of the classification decision arriving at that leaf node. In the context of a basic kNN classifier with majority voting, if all k nearest neighbours of a test image are from one class, then the classification decision will be most certain with the highest level of confidence. However, when there is a tie among the k nearest neighbours, i.e. the border line situation, the level of confidence regarding the class of the test image is the lowest. When the distances of the k nearest neighbours are considered, the level of confidence should reflect which class side the test data object should be more aligned to. In the context of an SVM classifier, the further away a test sample is from the separating hyperplane, the higher the level of confidence is about the class of the test image, and the nearer it is to the hyperplane, the lower the level of confidence becomes.

A measure of confidence in a decision made by a classifier will certainly depend on the criteria used by the adopted classifier. In this section, we shall attempt to introduce a simple, and easy to calculate, the measure of confidence together with a discrete concept of level of confidence that could be adopted in the area of automatic medical image diagnosis. We introduce a simple measure and levels of confidence for both kNN and SVM classifiers that is based on the distribution of distances/scores for test samples. Later in the chapter, we study the defined levels of confidence in the classification decisions made for the different gynaecological abnormalities. We shall close the chapter by focusing on multi-schemes fusion in general and expand this by introducing and testing the performance of confidence-based fusion for Ovarian Tumour identification.

6.3.2 Assigning Level of Confidence for the kNN Classifier

Consider the use of kNN as a classifier for a given binary classification problem, and assign the symbol “+” for one class and “-“ for the other class. With each classification decision, at

the testing stage, we associate one of three levels of confidence through the following method:

Step 1: Calculate the distance between the test sample and each template in the gallery, and normalise the set of signed distance measurements using the division-by-range method and then obtain the mean μ and the standard deviation σ over the distances. Then multiply each distance by the sign of the corresponding template class.

Step 2: Determine the k nearest neighbours of the testing sample.

Step 3: Calculate the average of the k nearest signed distance measurements and use the average as a decision score (Sc).

Step 4: Determine the level of confidence in the kNN decision by using Sc against the value σ, obtained in step 1, according to the following three rules:

Table 6. 1: Rules for assigning a level of confidence in kNN decisions

Rule	Condition	Level of Confidence
1	(Sc ≤ -σ) Or (Sc ≥σ)	**High**
2	(Sc ≥ σ/2 And Sc< σ) Or (Sc ≤ -σ/2 And Sc > -σ)	**Medium**
3	-σ/2 < Sc < (σ/2	**Low**

The rationale behind this confidence level setting scheme is that the overall mean distance defines a near zero separation of the positive and negative classes. The thresholds adopted in the rules are heuristically determined. In fact, assuming the right choice of gallery templates, the distribution of distances between any new sample and the gallery templates is expected to be normal. This means that there well-known rules about the fraction of samples' distances that fall within multiples of σ from the mean (e.g. there are around 68% and 95% of the scores within one and two standard deviations accordingly). The decision score, i.e. the average signed distances of the k nearest neighbours will give an indication on how far away the combined distance is from the mean and which side (+ or -) of it. The nearer the sample scores to the mean the less confidence one associates with the decision determined by the kNN rules.

6.3.3 Assigning Level of Confidence for the SVM Classifier

As explained in Chapter 3, section 3.5 and in (Duda, Hart et al. 2001), the SVM decision for a sample *x* depends on the score *Sc(x)* which represents the distance between *x* to the support

vectors near the hyperplane separating positive and negative classes. In this section, we first describe a set of rules to map a classifier score from an SVM classifier to an appropriate level of confidence.

It is natural to adopt the approach used for the kNN classifier for the SVM classifier and associate a decision confidence level to a sample x using the position of its score $Sc(x)$ within the distribution of decision scores of a training subset (a random sample) relative to the standard deviation σ of the training scores. Accordingly, we use the corresponding three rules of Table 6.1.

Figure 6.6 shows a hyperplane of an SVM with samples from the two classes being categorised into three levels of confidence based on their distance (i.e. scores) from the hyperplane. If the score of an image is on the far side of the positive samples from the hyperplane, then the sample will be classified as a benign with high level of confidence. However, if it's still on the negative side closer to the hyperplane, then it will still be classified as a benign but the level of confidence will be medium or low.

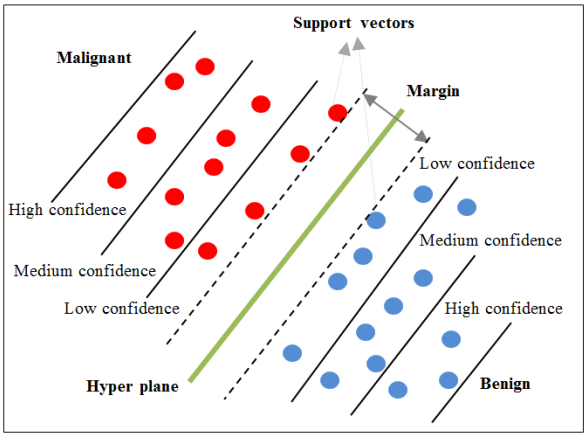


Figure 6. 6: Binary SVM with confidence level intervals

6.4 Confidence-related Accuracy for Gynaecological Abnormalities Classification

This section re-evaluates the accuracy of the main classification schemes, studied so far, within the level of confidence framework by implementing the above models of confidence measures. Based on the conclusions in section 6.2, we only workout the confidence levels using kNN only for the miscarriage identification and using SVM only for Ovarian Tumour classification. In each case, we use the best performing features.

6.4.1 Confidence Level of Miscarriage Classification Decision.

The identification of miscarriage using the kNN classifier with a number of geometric measurements of the scanned GS has been shown to achieve near perfect accuracy (about 99%) while the same or other features have significantly lower accuracy when classified with SVM. In this section, we investigate the effectiveness of applying the concept of confidence level to the miscarriage diagnosis problem, and for that, we revisit the distance calculations made to achieve the reported accuracy using the geometrical features of GS (such as MSD, Volume, and Perimeter) to calculate the level of confidence in discriminating miscarriage cases from PUV cases. Figure 6.7 show the block diagram for the process.

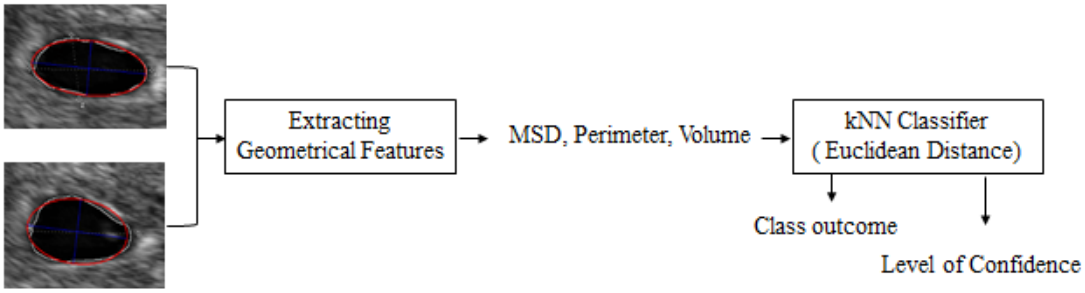


Figure 6. 7: The block diagram of confidence level computation of kNN decision

For these experiments, reported here, we set k = 3 for the kNN tests. Figure 6.8 presents the accuracy rate for the classification at the associated level of confidence for each type of feature first and then three types of features combined into a three component single feature vector. In other words, for each feature vector and at each level of confidence the displayed values represent the percentage of those images that are accurately classified at that level. For example, when the MSD feature is used in Experiment 1, then among the 90 test images, 39 images were classified with High level of confidence and 97.5% of them (i.e. 38 images) were correctly classified. Similarly, the decision on 40 other images were given the Medium level of confidence and 39= 40*0.975 image were accurately classified, while the decision about the remaining 11 images were given a Low level of confidence but only 9=11*0.8075 images were accurately classified. Similar innerpretation made about the 184 test images used in experiment 2.

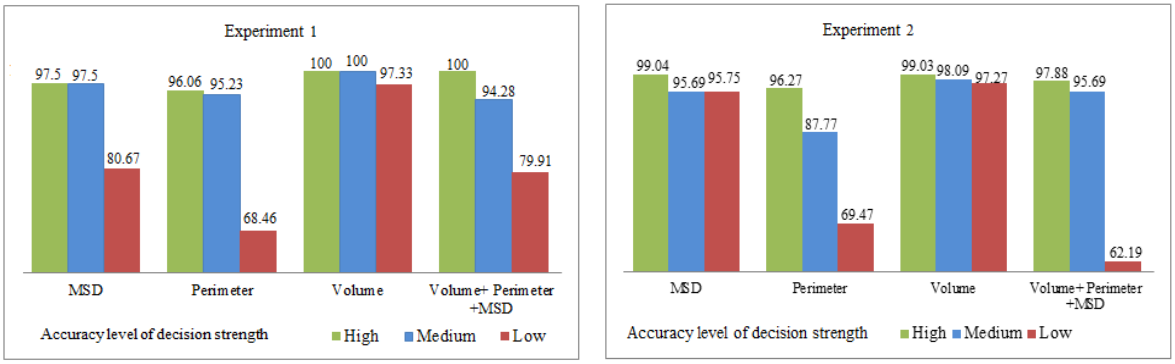
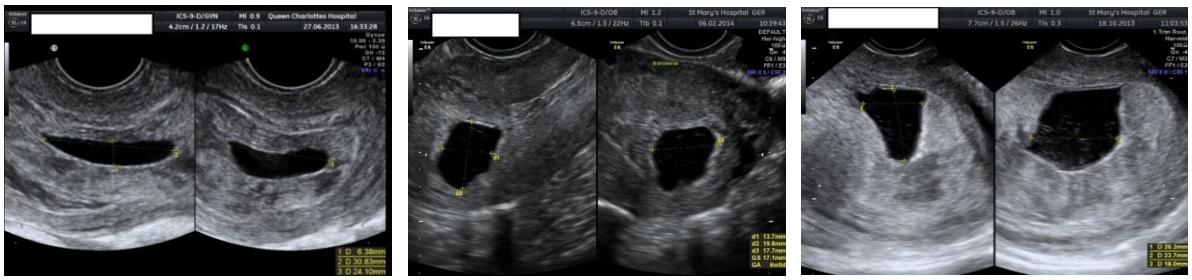


Figure 6. 8: Classification of miscarriage, for different features, at different levels of confidence

These results indicate a strong link between the level of confidence and accuracy rate for both experiements. In general, a decision made with high/medium level of confidence is highly likely to be accurate, while decisions made with low level of confidence could raise the alarm and in line with the medical practice it is safe to classify as PUV and a later scan should be conducted and re-investigated. A close examination of most images classified with low level of confidence are those where MSDs are measured between 13 mm and 26 mm, close to borderline cases (when MSD is between 16 mm and 25 mm) according to the guideline given in (Liao, Wan et al. 2011). Figure 6.9 show some examples of such images. The borderline cases in practice are more challenging to diagnose even for specialists. We note that there is a certain degree of disagreement between the United States guideline (MSD > 16mm) and the UK guideline (MSD > 25 mm). This means that our decision strength levels correspond closely to the guideline's findings.

Furthermore, we also have found that most images that classified with the High level of confidence are of miscarriage cases whereas most images that classified with Low or Medium levels of confidence are PUV cases. This finding is also somehow expected: it is more certain to diagnose large size clear empty gestational sacs as miscarriage cases than identify various complexes and often less certain pregnancy of uncertain viability (PUV) cases. More work is needed to look into the factors that affect the outcomes, and ultimately increase the certainty element of PUV cases.



(a) (b) (c)

Figure 6. 9: Examples of PUV cases with Low level because they are much near to border line (a) MSD=20.385 (b) MSD=17.897 (c) MSD=24.071

6.4.2 Confidence Level of Ovarian Tumour Classification Decision.

The classification of Ovarian Tumours (into Benign and Malignant) using the SVM classifier with the various image-texture based features extracted from Ovarian US images have been shown to significantly outperform their classification by the kNN. In Chapter 5, we found that FFGF, LBP (256) and Histogram were the best three performing features when Ovarian Tumour US images were classified by SVM. Figure 6.1 includes the accuracy rates achieved by each of these three single features within each level of confidence when classified by SVM. These rates are calculated as the average from 15 repetitions of the experiment rounds, and each represents the average percentage of accurately classified images among those given the corresponding level of confidence. For example, when FFGF was used 22 images (out of total input of 100) where classified with High level of confidence of which 19.2=22x0.8638 images were correctly classified, 38 images were classified with Medium level of confidence of which 35.33=38x0.9296 images were correctly classified, and the other 40 images were classified with Low level of confidence of which 30.6=40x0.7687 images were correctly classified. Note that, the LBP has significantly lower overall accuracy than the FFGF but has higher accuracy at the high level of confidence.

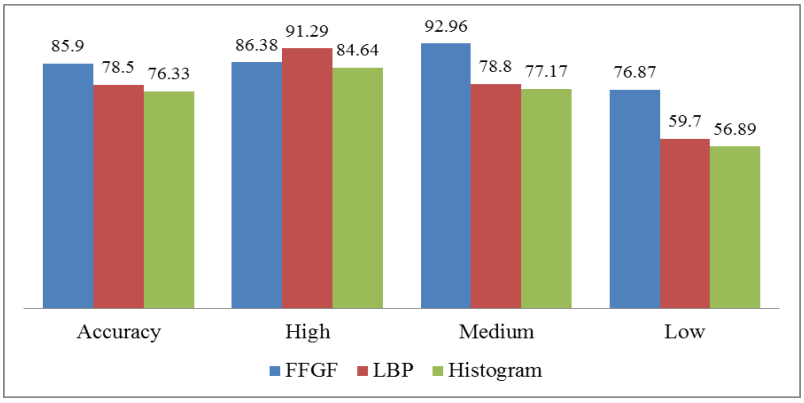


Figure 6. 10: SVM Classification of Ovarian Tumour image, for different features & confidence levels

6.5 Fusion of Multi-Schemes Classification of Ovarian Tumour

The results of using single classifier for distinguishing benign and malignant ovarian tumours with the various image-texture based feature vectors have revealed a range of reasonable to high degree of accuracy especially by SVM, but this still well below the desired accuracy for such a life-threatening disease. As mentioned earlier, the fusion of multi-schemes pattern

recognition is a common practice for improved accuracy. At the heart of our proposed fusion framework, there are two fundamental ideas: (a) single classifier fusion of multi-feature vectors schemes and (b) fusion of multi-classifiers that use the same or different feature vectors. Here, we shall confine our investigations to the first approach and, whenever possible, the incorporate level of confidence associated with each of the different feature-based schemes.

The single classifier multi-schemes fusion is a possible way of improving accuracy as long as there are a number of different feature vectors representing the space of data items, and a classifier exist with which each of these feature vectors define a reasonably recognition scheme. In Chapter 5, we established that a variety of image-texture feature vector representations of US scan images of Ovarian Tumours that lead to considerably successful classification when fed into a binary linear SVM classifier. But the performance of these schemes has reduced when fed into the kNN classifier. In this section we investigate various applicable fusion models. Here we investigate different fusion models that we described earlier in Chapter 3.

6.5.1 Feature Level Fusion

In this model, two or more feature vectors of the data items (i.e. US images) are concatenated to form one feature vector before input into a chosen classifier. Taking into account the three best performing image-texture based feature vectors we repeated the experiments conducted in Chapter 5 and in section 6.1, above, we tested the performance of various schemes that can be formed by the combinations of the three identified effective feature sets, i.e. Blocked based histogram (256 bins), Blocked based histogram of (LBP 256 bins) images and FFGF (Major, Minor and Area). Figure 6.11 depicts the workflow of the feature-level fusion, whereby we apply the SVM classifier to each combination of the three feature vectors.

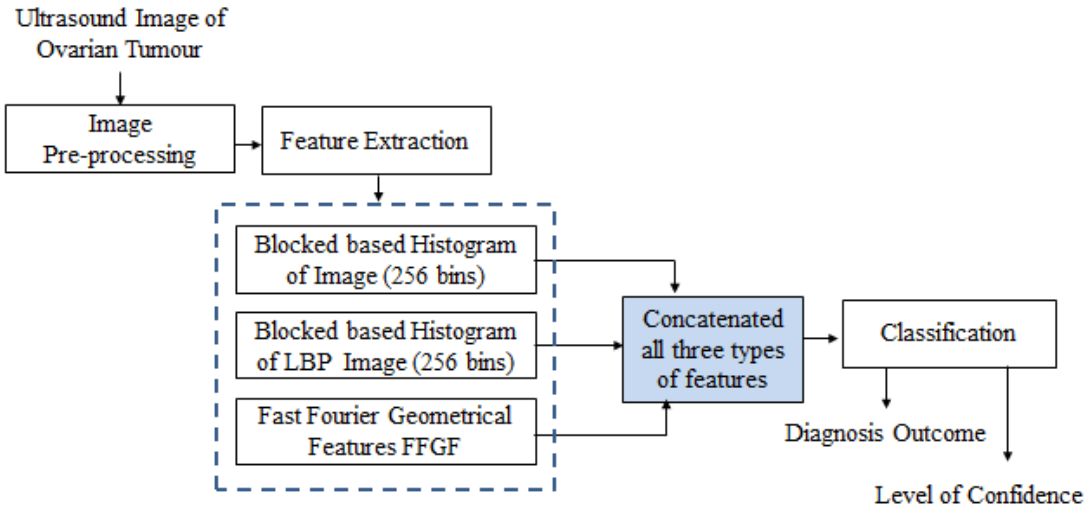


Figure 6. 11: Flow diagram of the feature level fusion

Figure 6.12 shows the overall accuracy rate achieved when SVM was used with each meta-feature obtained by concatenating different combination of original features, together with the associated accuracy rates within each confidence level. For comparison we also include accuracy rates for the individual features. The results show that combining the histogram with the LBP features leads to a similar level of accuracy of the LBP feature alone, and hence the histogram has no significant "added value" on the LBP feature although the improvement over the histogram feature alone is still nearly 4%. However, combining the histogram, the LBP or both with FFGF produce worse results than using the FFGF only. These results show that feature based fusion does not improve accuracy relative to the best performing features but only relative to the less performing component(s). This is possibly due to the huge difference in the dimension and the nature of the vector components for the histogram or the LBP and the FFGF features. For instance, in the combined LBP+FFGF meta-feature vector, there are 4x4x256 attributes for LBP, and yet only three attributes for FFGF. This extreme difference means the influence of FFGF in the combined meta-feature vector is very small. Comparing the accuracy rates within each level of confidence, we can see a slight

improvement

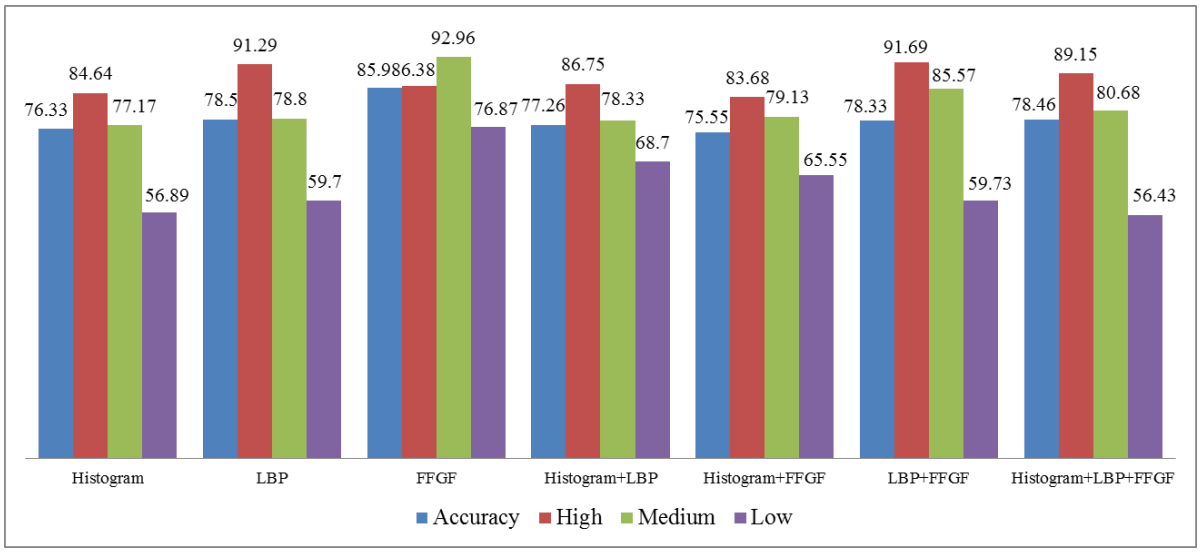


Figure 6. 12: Performance of SVM for the feature level fusion with levels of confidence

We conducted the same experiments for feature level fusion but using the kNN, and the results are presented in Figure 6.13, below. As expected for all feature level fusion schemes, the performance of the kNN is significantly outperformed by the SVM versions. Moreover, just as in the case of SVM, with the exception of the FFGF+LBP scheme, each feature level fusion with kNN is outperformed by single feature schemes on their owns.

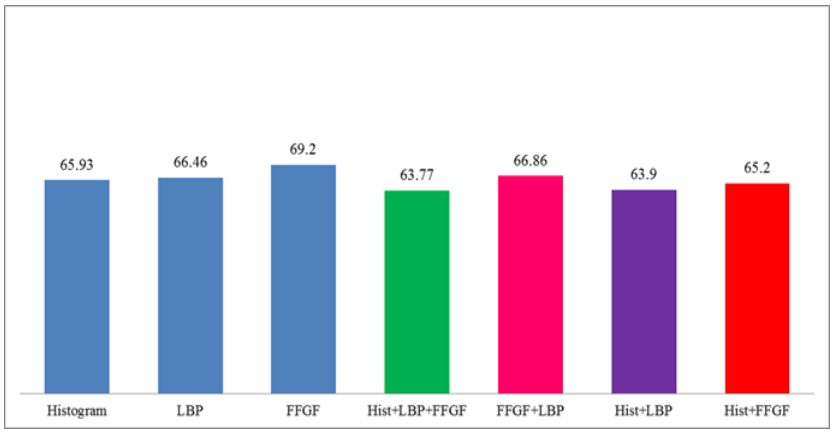


Figure 6. 13: Performance of kNN for each feature alone and concatenated features

The rather disappointing failure of fusion at the feature level is most likely to be the differences in the nature of the component feature vectors.

6.5.2 Score Level Fusion

The next common model of fusion, whereby the classification scores for all the individual feature scheme are combined according to some rule to deduce aa fused score that will be used for final classification decision. We fuse at the score level various combinations for the

same three individual feature vectors (Histogram, LBP and FFGP) and also determine the levels of confidence for each combination using the SVM related rules described in Section 6.3.2. First, the SVM classifier is applied in each feature space separately, and records the signed score for each feature. We then take the average of the three signed classification scores obtained for the three features. The sign of the average score determines the final class, and the average of score value is then mapped to a level of confidence according to the rules presented in Section 6.3.2. Figure 6.14 illustrates the workflow of the fusion method.

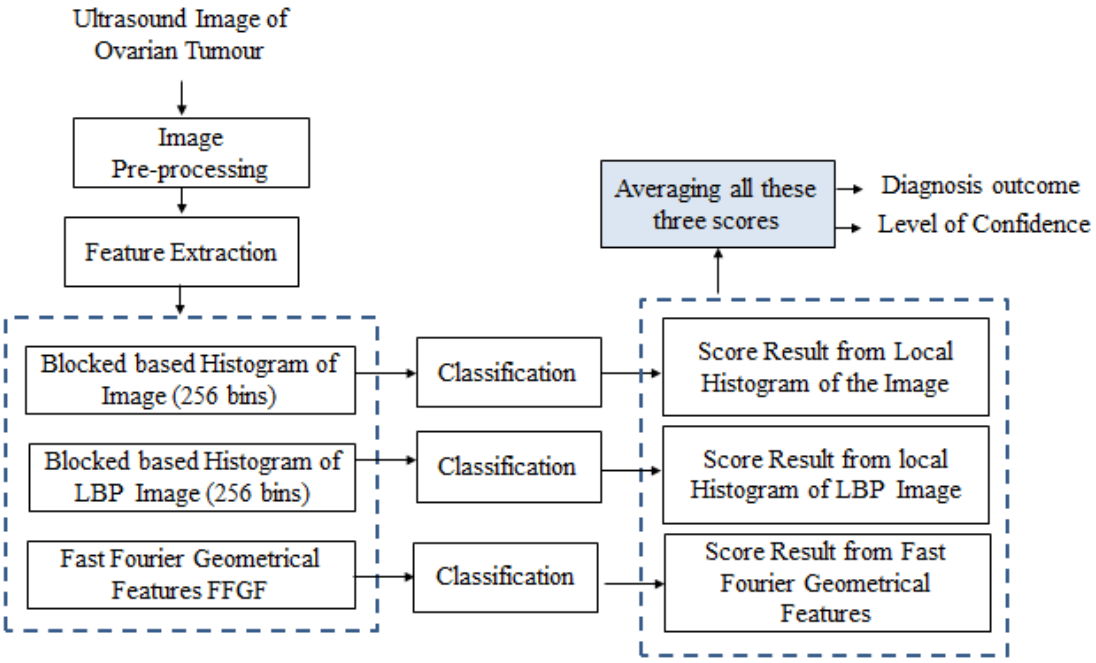


Figure 6. 14: Block diagram of the score level fusion

Figure 6.15 presents the classification results of the score level fusion of the various different combinations of the three image texture-based features, by listing the overall accuracy rate with and without the reference to the three levels of confidence. Recall that the rate reported for any fused scheme at each of the level of confidence represent the percentage of US tumour images that are correctly classified out of all images for which the classification decision were assigned the given confidence level. For example, the FFGF+LBP score fusion scheme achieves an overall accuracy rate of 88.53, and for every 100 testing images that are classified at the High level of confidence 93.74 of these images would have been accurately classified.

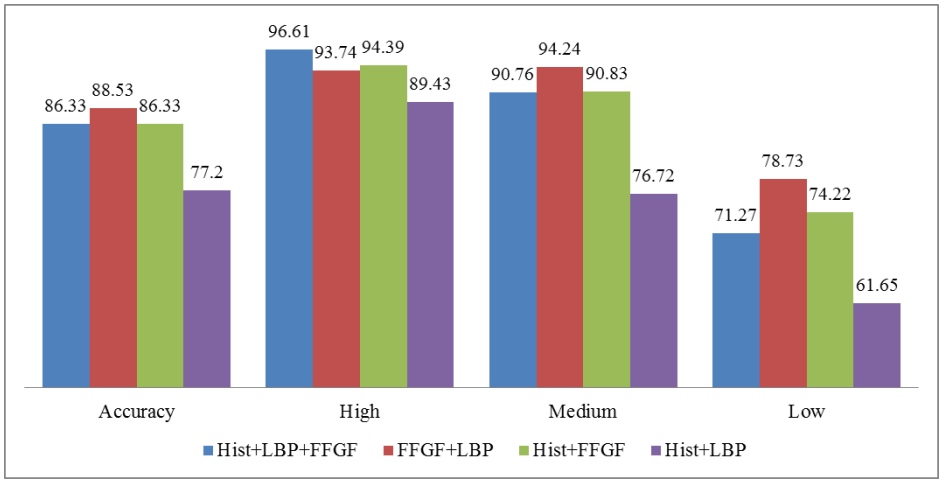


Figure 6. 15: SVM Score level fusion with for the 3 levels of confidence

The results reported above demonstrate that the best accuracy (88.53%) is achieved when scores of FFGF is fused with scores of LBP representing an absolute 2.2% increase on the original FFGF scheme. These two features complement each other in capturing the image texture, one at the local level in the spatial domain (LBP) and the other at the global frequency domain (FFGF). The histogram does not contain texture information explicitly. Out of the three levels of confidence, the FFGF+LBP combined scores also outweigh significantly the combined scores of three types of features for the medium and low levels confidence but not at the high confidence level. Interestingly comparing fusion at the score level with fusion at the feature level, for the various combined schemes, reveals that only for the combination of Histogram and LBP features the score level and the feature level fusions have similar accuracy, but in all other cases the score level fusion significantly outperforms the feature level fusion.

Replacing the SVM classifier, in the simple score level fusion experiments, with the kNN classifier has revealed an opposite pattern of performance to the SVM results. Figure 6.16 confirms that all the score based fused schemes are significantly outperformed by each of their single feature schemes (see Figure 6.1). Not only, the SVM score fusion schemes outperform the kNN ones but the confidence measure for each of the three levels are significantly lower.

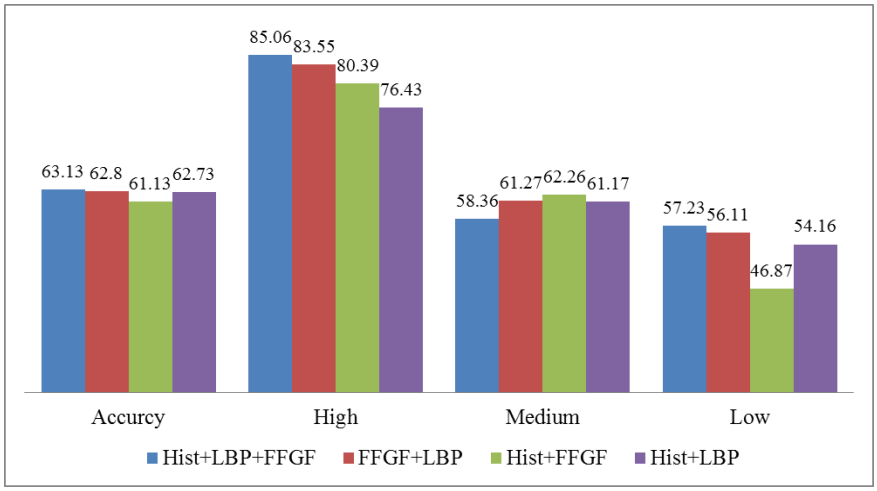


Figure 6. 16: kNN Score level fusion of the 3 levels of confidence

The fact that there significant differences between the performance of each of the three individual features (85.9% for FFGF, 78.5 for LBP, and 76.33 for Histogram) is a motivation to apply the alternative form of score level fusion known as the *Weighted Average Score level fusion.* This method follows exactly the same workflow of score level fusion, accept that the various individual scores are added with different weights that reflect their individual classification performances as reported above, i.e. we use the classification accuracy rate of each type of feature divided by the total accuracy rate of the three types of features. Since 76%, 78%, and 86% are respectively the reported accuracy rate for the Histogram, the LBP and the FFGF features, the classification score weight for each feature is to be set to 76/(76+78+86), 78/(76+78+86), and 86/(76+78+86), respectively.

We are slightly modified score level fusion block diagram shown in Figure 6.14 to reflect the workflow for this method by incorporating the weighted average with the above weights in the last step of computing the score level fusion. We then repeated the same experiments above with this strategy using the SVM classifier. Figure 6.17 presents the achieved accuracy rates together with the confidence rate for each of the three levels of confidence. The accuracy rates for each of the schemes has increased but by a marginal amount < 0.93%.

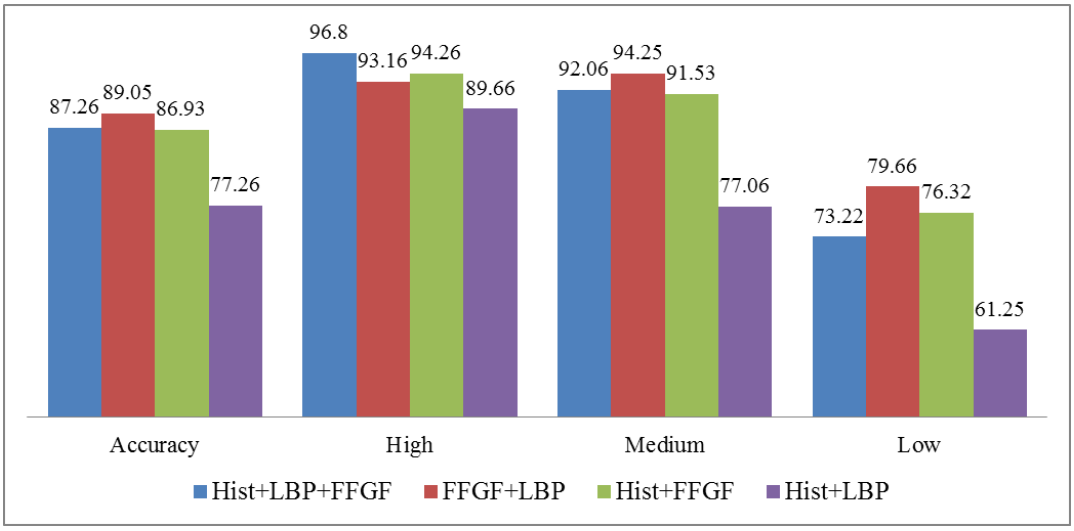


Figure 6. 17: Weighted score level fusion with the confident level based

These results demonstrate modest benefits, in terms of accuracy, of using the weighted average instead of simple average based score fusion. However, by looking at the three levels of confidence, the confidence rates are marginally reduced for the high and medium levels, but increased for the low confidence level has been increased by 1.95%. Therefore, the increase in accuracy is achieved at the expense of increased border cases that would require further medical intervention. This performance pattern seems to indicate that the weighted score fusion based on the individual feature performances is crude and has a limited success, and perhaps one can find a slightly modified weights to avoid the disadvantage of losing confidence rate. However, this would require the extensive set of experiments together with careful consideration with little benefits for the current binary classification of Ovarian Tumours (Benign Vs Malignant).

Replacing the SVM classifier with the kNN classifier was not expected to change the emerging pattern of results. But we still decided to conduct the weighted-average score level fusion schemes of the various combinations of three feature vectors for the kNN classifier, to see if this provides a more positive impression that has emerged so far about the performance of the various fusion approaches when kNN is used. Here, the final weighted average score was calculated in a similar way to the SVM case when the weights was meant to reflect the relative performance of each feature, i.e. the final score is obtained by the following formula:

Score (FFGF)*0.343+ Score (LBP) * 0.33+ Score (Histogram)* 0.327

Figure 6.18 demonstrates that the accuracy rate for each weighted average score fused scheme is identical with that achieved by the corresponding simple score fused scheme. The

confidence measures across all levels are subject to some changes for few schemes in a manner that could not lead to improving the emerging impression of kNN being irrelevant to the classification of Ovarian Tumour from US images.

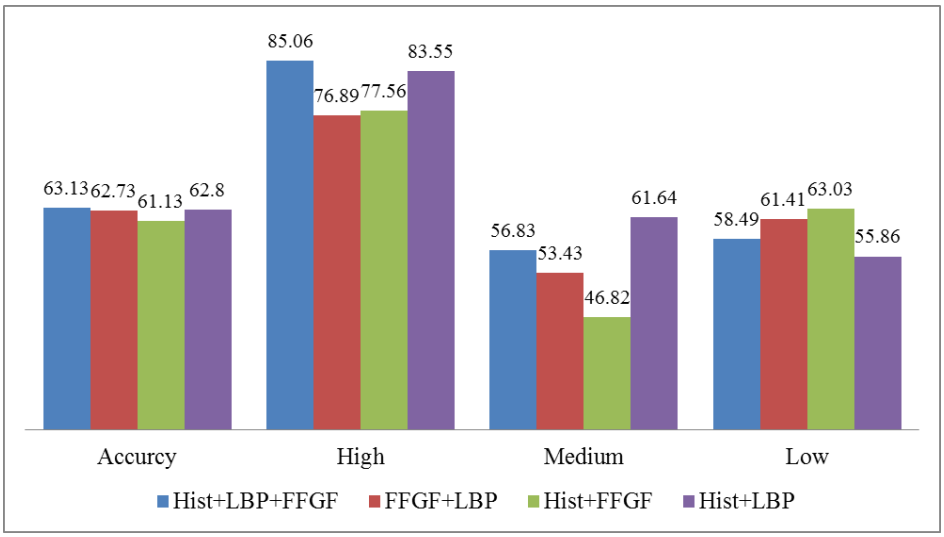


Figure 6. 18: Score level fusion with for the three levels of confidence

The similarity in the performance of weighted average score level fusion to that of the simple score level fusion for the kNN classifiers is most influenced by the chosen weights which are almost equal reflecting the fact that all the three features achieve nearly the same accuracy with kNN. However, we repeated the same experiments with a set of different weights (0.42 for FFGF, 0.32 for LBP and 0.26 for Histogram) the accuracy rates of the fused scheme did not change in a big way though was lower than in the above table.

6.5.3 Decision Level Fusion

This fusion model differs from the score level fusion in that instead of just using the score for each feature scheme, the final decision will be based on the decisions made for each feature component scheme as a voting system. The simplest and most common criteria for making the final decision would be based on the class that receives the largest number of vote. Since we are dealing with a binary classification problem then fusing an odd number of feature schemes then the majority rule will always yield a clear decision. We have implemented the fusion of the three image-texture based schemes (Histogram, LBP and FFGF) at the decision level, and tested the performance of this scheme using the same experimental setup in the previous section for simple score-level fusion, but for comparison, we used both classifiers SVM and kNN. Figure 6.19 presents the results of these experiments.

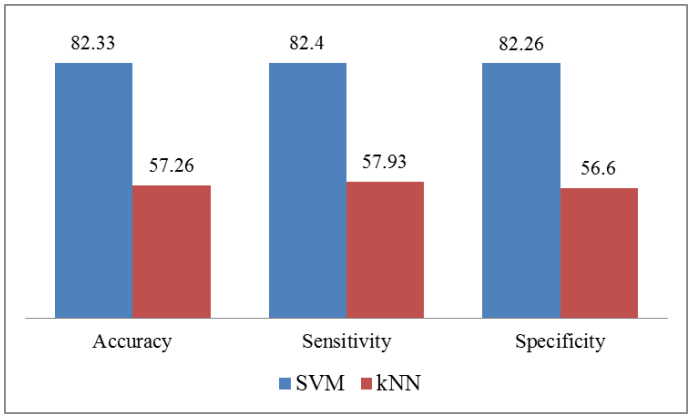


Figure 6. 19: Decision based Fusion –Majority rule

These results demonstrate that the majority rule for Decision based Fusion is outperformed considerably by the single best-performing feature scheme (FFGF) for the SVM classifier, but for the kNN classifier it is significantly outperformed by each of the components features (see Figure 6.1). Nevertheless, in the case of SVM, the Decision based Fusion outperforms the corresponding feature level fusion. Moreover, the both types of score level fusion of the same three feature schemes significantly outperform the majority Decision Fusion.

The feature level fusion and score level fusion have both improved accuracy when two feature components rather than the three features are fused. Hence, the rest of the chapter is devoted to investigating Decision Level Fusion for each pair of the three image-texture based features. The problem under investigation, being a binary classification, presents a challenge when only two feature schemes are available due to the fact that the majority voting decision criterion is no longer possible. When only two classification decisions are to be fused the challenge is what to do when the two schemes disagree. Our approach to deal with this problem is to take into account the confidence measure/level associated with the two decisions. Classifying a sample simply by the class of higher confidence, among the two different decision, lead to another question as to what would be the confidence in the fused decision? The last question also arises even when both decisions identify the same class. In the rest of this section and chapter, we attempt to answer these questions and develop a decision level fusion model that yields classification outcomes supported by the level of confidence. We also develop rule for associating the level of confidence to the output classification decision. In the proposed scheme, only one classifier (the SVM) is used for reasons explained throughout the chapter about the kNN. Moreover, only pair-wise fusion of the classification outcomes from two features is considered. Figure 6.20 outlines the framework of the proposed fusion scheme.

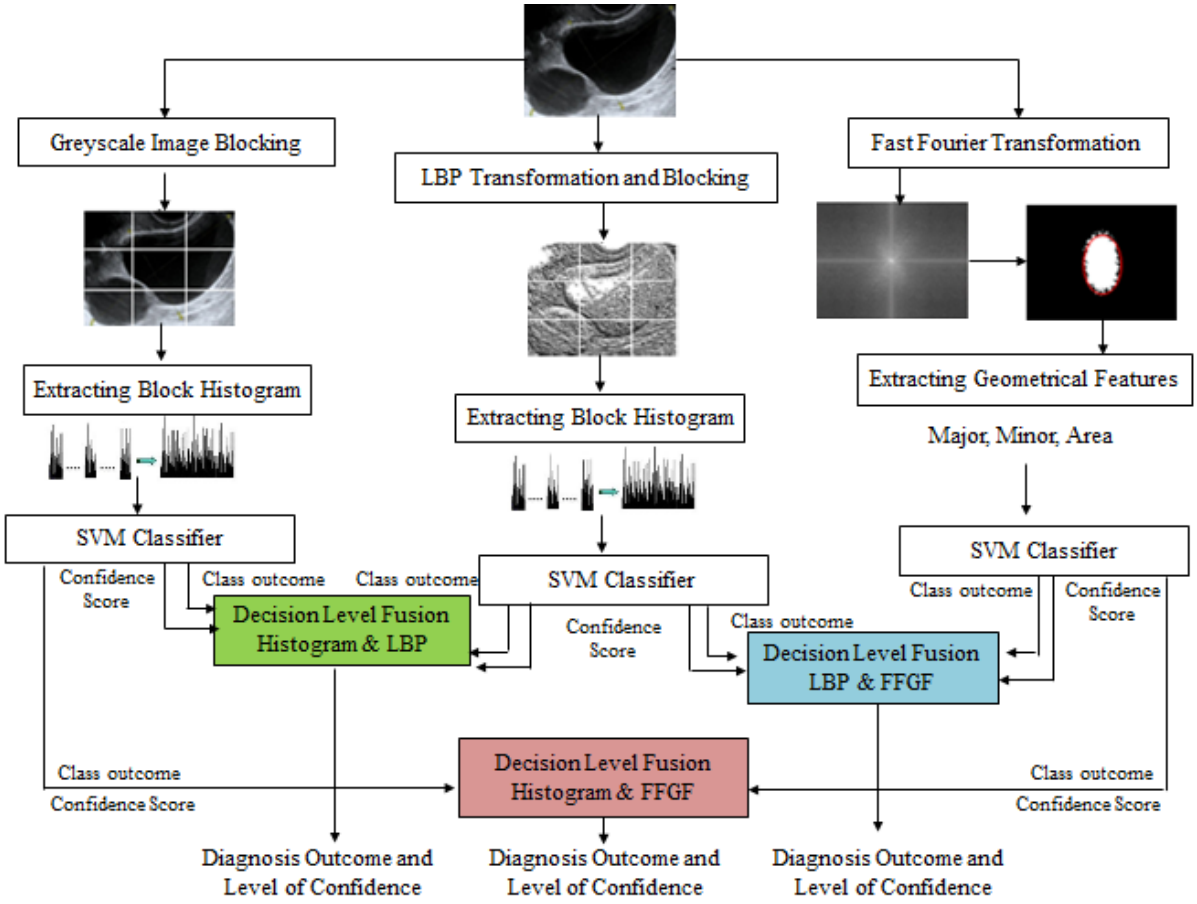


Figure 6. 20: SVM based Decision level fusion of pairs of feature schemes (Histogram, LBP, & FFGF)

The proposed fusion scheme initiate the SVM classification of an input sample and feeds into the decision level fusion unit the classes chosen by the single feature schemes together with the corresponding computed confidence level. The fusion unit outputs a final class decision and determines the level of confidence of this decision using a specially designed criterion, described below. Recall that for any SVM classification scheme we already have three rules that governs the position of the classification score the level of confidence as presented in Table 6.1, according to the final classification outcome by the fusion, also requires a clear set of rules to define the level of confidence of the outcome. The rules should cover two possible scenarios: (a) when the pair of classifiers yields the same class outcome, and (b) when the pair of classifiers yields two different class outcomes. Table 6.3 presents the definition of the level of "combined" confidence based on the level of confidence from each of the two SVM classifiers for all possible situations. The rules can be explained in this table below:

Table 6. 2: The main rules of decision level fusion scheme

Rule 1
If the class outcomes are the same and the levels of confidence are the same before fusion, then the fused class outcome is the same class, and the level of "combined" confidence remains the same.
Rule 2
If the class outcomes are the same but the levels of confidence are different before fusion, then the fused class outcome is the same class, and the level of "combined" confidence is adjusted according to the specific values as follows: o If one level of confidence is High and the other is Medium before fusion, the level of "combined" confidence is set to ***High***; o If one level of confidence is High and the other is Low before fusion, the level of "combined" confidence is set to ***Medium*** (i.e. the middle value); o If one level of confidence is Medium and the other is Low before fusion, the level of "combined" confidence is set to ***Low***.
Rule 3
If the class outcomes are different and one level of confidence is High and the other is Low before fusion, then the fused class outcome is the class with High level of confidence before fusion, and the level of "combined" confidence is set to Low.
Rule 4
In any other cases, the fused class outcome is set to ***Not Sure***, and the level of "combined" confidence is set to ***unspecified***.

The rationale behind these rules can be summarised as follows:

1. When the class outcomes from both classifiers are the same, it means the classifiers agree with the outcome. The fused level of confidence should reflect the true strength of confidence in the decisions. When one prior level is High, and the other is Low, setting the fused level confidence to Medium is a compromise. Setting the combined confidence to High, when one prior confidence is High, and the other is Medium, it is prudent to set the combined confidence to Medium. We also take a more cautious stand and set the combined confidence to Low when one prior level of confidence is Medium and the other is Low. This approach is partially influenced by the views of end users.

2. When the two classifiers yield different class outcomes, the fused decision needs to resolve the conflict. Therefore, the prior levels of confidence become important in the final judgement. If one class prediction is made with High confidence and the other class with Low confidence, the decision with High confidence should win in determining the

final class outcome, but the combined confidence should reflect a reduction in the strength of the class prediction. Again, we take a quite prudent stand on this and decide to set the level of combined confidence to Low in this case to reflect the conflict in the class outcomes.

3. We introduce a special case of the *Not Sure (NS)* class to reflect strong opposite class predictions from the two classifiers. In total, this may happen in 14 cases. Table 6.3 below depicts these rules in a concise manner.

Table 6. 3: Rules for Decision Level Fusion

Case 1 / Case 2	HB	MB	LB	HM	MM	LM
HB	HB	HB	MB	**NS**	**NS**	LB
MB	HB	MB	LB	**NS**	**NS**	**NS**
LB	MB	LB	LB	LM	**NS**	**NS**
HM	**NS**	**NS**	LM	HM	HM	MM
MM	**NS**	**NS**	**NS**	HM	MM	LM
LM	LB	**NS**	**NS**	MM	LM	LM

*Here, **HB**: High Benign, **MB**: Medium Benign, **LB**: Low Benign, **HM**: High Malignant, **MM**: Medium Malignant, **LM**: Low Malignant, **NS**: Not sure*

The over-prudent decision of classifying the 14 NS cases is a reflection of the fact that in medical applications such automated decisions should be subject to medical expert's advice in order to avoid serious consequences of errors in classification. Perhaps one could relax the rules slightly to reduce the number of NS cases, but this kind of investigations will be left for the future. Although, the introduction of the NS class is desirable it raises a question on how to rate these cases when one evaluate the performance of the classification scheme: Should the declared NS cases be considered as success? Or should these be considered as misclassification? Moreover, since this concept is not relevant to all other schemes discussed above, then it is difficult to compare the performance of this with that of the earlier ones. In practice, medical experts often ask for the second opinion or even request more tests/procedures to be conducted when they have no convincing evidence either way, and such an approach by the expert will not be seen as misclassification. In other words, declaring NS in automatic diagnosis should not be considered as a failure, but should be taken as good

practice although it would be difficult to consider all NS cases as successful classification. In what follows, we present performances of the confidence based Decision level fusion for different scenarios.

We conducted a new set of experiments with the same dataset of US scans of Ovarian Tumours using the above rules in the context of the decision fusion scheme. We follow the evaluation protocol as described in Chapter 3, Section 3.7. Taking into account the above discussion regarding the NS cases, we first present the accuracy rates by excluding the NS cases from the calculation of the accuracy rates. Figure 6.21 shows the decision level fusion accuracy with this scenario together with "combined" confidence levels of high, medium, and low. We conclude that when we ignore the NS cases the decision level fusion significantly improved the classification accuracy in comparison with using each feature set alone, feature level fusion and score level fusion of the corresponding pair of features. For example, the decision level fusion of FFGF+LBP achieves a high accuracy of 93.87% against 78.33% for the corresponding feature level fusion, 88.53% for score level fusion, and 89% for the weighted score level fusion. Moreover, the accuracy achieved within the different levels of confidences is significantly higher than the weighted and unweighted score level fusion schemes.

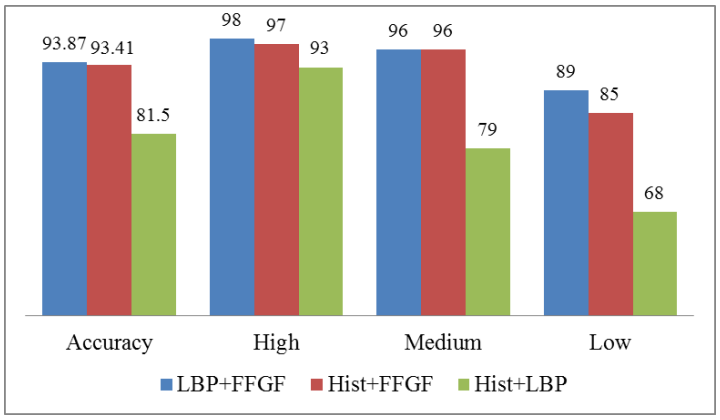


Figure 6. 21: Accuracy of decision level fusion with confident based level

For the above experiments, the average numbers of NS cases were as follows: 25 cases for (Hist+ FFGF), 18 cases for (Hist+LBP) and 22 cases for (LBP+FFGF). Note that the averages for NS and the accuracy are worked out over 15 repetitions of the experiments. Table 6.4 below gives, these averages of NS cases for the three decision fused schemes together with corresponding averages for Benign and Malignant classes.

Table 6. 4: Average No. of NS cases for the three decision fused schemes			
Decision Fused Scheme	**All**	**Benign**	**Malignant**
FFGF + LBP	22.6	11.7	10.9
FFGF + Histogram	25.4	12.8	12.6
LBP + Histogram	17.6	9.9	7.6

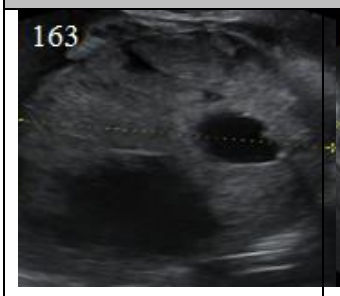
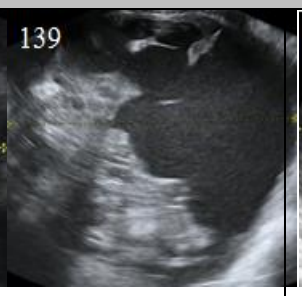
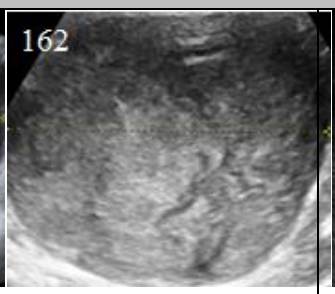
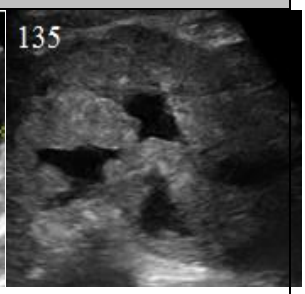
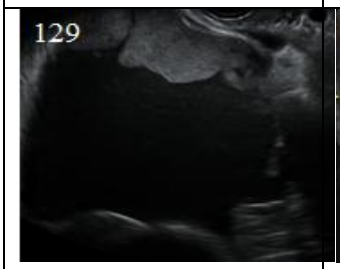
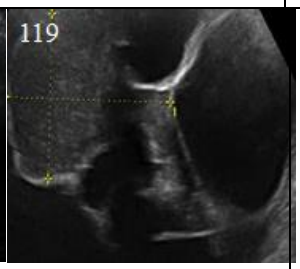
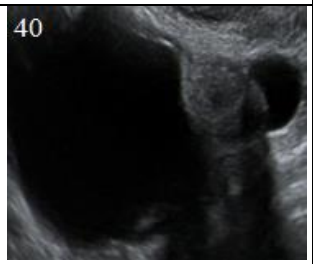
A prominent collaborating expert in the field commented that most of our NS cases could be better classified as Low malignant due to a rule of thumb practised in real-life clinics "*if in any doubt, consider the tumour as malignant*". If we to heed this advice and declare all NS cases as Malignant (nominally with Low confidence) then the estimated average accuracy for the three schemes will be reduced to: 83.6 % for (FFGF + LBP); 82.2% for (FFGF + Histogram); and 74.8% for (LBP + Histogram). These results are certainly lower than those in figure 6.20 where we disregarded the NS cases. The decision level fusion with majority rule of the three features FFGF+ LBP+ Histogram is slightly outperformed by the FFGF+LBP scheme, and only marginally better than the FFGF+ Histogram (see Figure 6.19).

To shed more light on the above set of different results and analyse various data, we retrieved and examined the sets of NS images for each of the three fused methods. Interestingly, different sets of images were labelled as NS in the different 15 repetition rounds of the experiments. However, close examination of those lists revealed that the seven images are shared in all 15 rounds and form the intersection of the NS image sets in the three fusion methods. We found that images that been considered as NS cases most of them were malignant. Furthermore, we discussed with experts the classification of those seven NS images and they confirmed that those images are very difficult to classify as well by the expert. These images are displayed below in Table 6.5. Interestingly, six of these seven images are malignant and this is in line with the expert advice given earlier.

Further extensive work is needed if we to incorporate expert advice into the decision level fusion to deal with NS cases. Such work should aim at reducing the number of NS cases by involving other image-based features and/or modifying the levels of confidence intervals. One approach to reducing the number of NS cases might involve subdivided the current confidence levels into sublevels so that to facilitate the resolving many more conflicting classifications. Alternatively, is to change the boundaries using different multiples of the standard deviations.

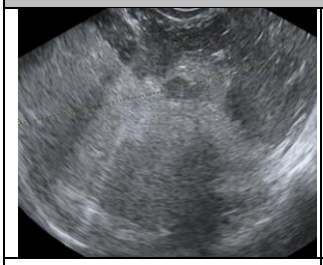
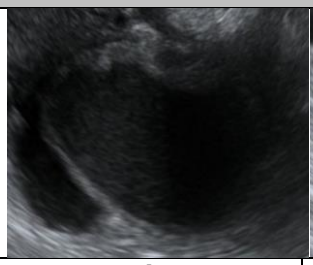
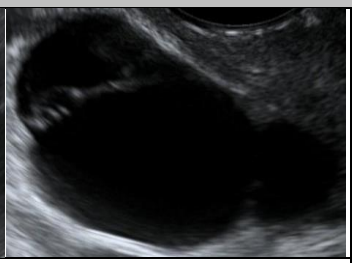
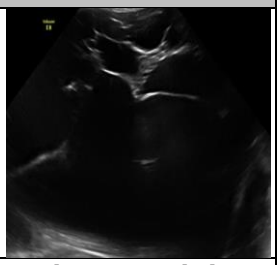
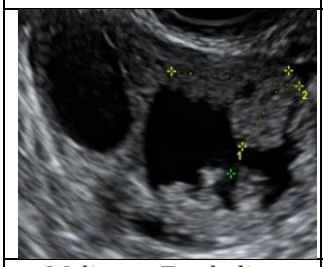
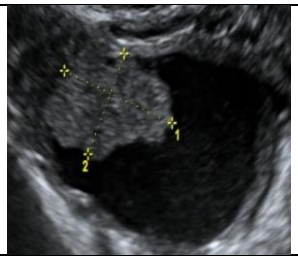
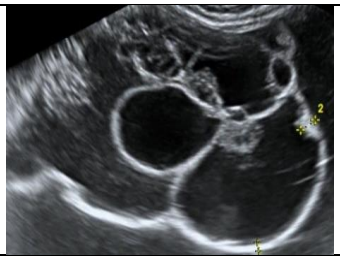
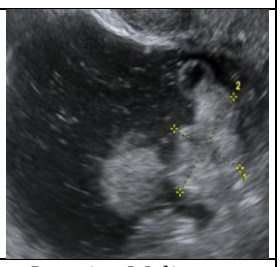
Finally, we remark that table 6.3 is generic and doesn't take into account any knowledge that we may have about the performance of each of the contributing scheme at each level of confidence. For example, we found in Figure 6.10 that at the High level of confidence, the LBP has higher accuracy than the FFGF, this may justify reducing the cases of NS when we have LBP returns a HM decision while the FFGF returns a MB decision especially if the corresponding scores are well separated. However, such consideration requires more investigation using the larger number of samples.

Table 6. 5: The seven NS images common to all three decision level fusion schemes

			
Invasive malignant	Invasive malignant	Invasive malignant	Invasive malignant
			
Borderline	Borderline	Serous cystadenoma	

We close this section by showing several images that were misclassified by each feature set alone but became NS by the confidence based decision level fusion method. These examples are shown in Table 6.6 and illustrate the benefit of introducing the NS case, as misclassification could have serious health consequences.

Table 6. 6: NS images that were originally misclassified by the various single features

			
Benign/ Teratoma	Benign/Endometrioma	Benign/Serouscystdenoma	Malignant/Borderline
			
Malignant/Borderline	Malignant/Borderline	Metatastic Malignant	Invasive Malignant

6.6 Summary and Conclusions

This chapter builds on the work done and a conclusion made in the last chapter, and investigates various aspects of machine learning approach to consolidate and/or improve the performance of the automatic binary classification schemes that introduced in Chapter 5, for the different applications of miscarriage detection and ovarian tumour identification.

We first extended the performance testing investigations, conducted in Chapter 5, in terms of discriminating power of the different sets of application-dependent image feature vectors using alternative classifiers. The experimental works confirmed that the SVM classifier significantly outperforms the kNN classifier when used with each of the image-texture based features (FFGF, LBP, and Histogram) for distinguishing benign and malignant ovarian tumours. However, the experiments demonstrated that kNN outperforms the SVM but not by a relatively small margin when used with each of the GS measurement features (MSD, Perimeter, Volume, etc.) to identify miscarriages from US scan images GS. The last set of results and the fact that for Ovarian Tumour classification the SVM yielded the best accuracy with the low-dimensional FFGF feature suggests that the common wisdom that SVM is best suited for high-dimensional problems may not be valid for every type of low-dimensional features, especially those that encapsulate the entire application-relevant information hidden in a high-dimensional space.

We then turned our attention to the goodness of automatic decision making by introducing a quantitative measure and level of confidence in a decision relevant to the two applications

under consideration in this thesis. This was motivated by the need to emulate what happens in real life medicine when diagnosing health related diseases from medical radiology images. When well-experienced specialists made final decisions they use well-established sets of rules and backed up medical evidence, but when the diagnosis is inconclusive further medical checks such as surgery/biopsy on the patient are conducted. The results of various experiments indicate a strong link between the level of confidence and accuracy rates so that a decision with the high/medium level of confidence is highly likely to be accurate while decisions made with the low level of confidence could raise alarm in line with medical practice.

Investigating the various approached to fusions of multi-feature schemes revealed some interesting challenges. First of all, feature level fusion cannot improve on any of the single components perhaps due to differences in the nature of the various participating features. Score level feature fusions (weighted/unweighted average) have led to improved results but there is more potential for improving the performance of classification of Ovarian Tumour cases. The decision level fusion of three sets of features using majority rule outperformed the feature-level fusion scheme but was still less than the best single feature. On the other hand, our pilot innovative decision level fusion of two sets of features supported by the level of confidence measures provided a chance to make a Not Sure decision and with some positive consideration of this one can get a significantly improved accuracy. Again, extensive future research is needed to develop this approach to a mature system that can be part of a wider Decision Support System in this field. In the next chapter, some light will be shed on possible ideas along this line.

CHAPTER 7

CONCLUSION AND FUTURE AND ONGOING WORK

Since the introduction of X-ray into medical practices early last century, there has been growing development and interest in the deployment of a variety of medical imaging systems for diagnostic purposes. This thesis was devoted to investigate and develop automatic tools for computer-aided diagnosis of gynaecological abnormality from B-mode ultrasound images. Accurate and early diagnosis is essential in the management of gynaecological problems such as ovarian tumour and miscarriage diagnosis. It facilitates effective treatment and prevents developing into serious stages of the disease. For example, distinguishing benign from malignant ovarian tumours is important, not only to ensure proper treatment for tumour types, but also to avoid unnecessary diagnostic procedures such as biopsy/surgery. However, early detection of ovarian carcinoma continues to be a difficult challenging task not only due to technical problems but due to factors relating to patients unawareness of relevant symptoms as well as limited availability of well trained staff to deal with the volume of data and US scan images. On the other hand, the first trimesters are the most crucial period in pregnancy to evaluate the development, growth, and wellbeing of the fetus and the mother. The first sign and measurable element of early pregnancy is the appearance of the GS. Quantifying the GS shape and dimensions from US scan images is the common medical tool for identifying the cases of miscarriage. Early detection of miscarriage is essential in preventing health complications for mothers.

Our investigations were motivated by recent successful deployment of image-based pattern recognition and datamining techniques on one hand and the growing realisation that machine learning is becoming necessary to deal with rapidly growing demands in the health service. This work was conducted in collaboration with UK and European prominent medical experts as part of an effort to eventually develop a decision support system in this field. The rest of the chapter is organised as follows: Section 7.1 summarizes the research work carried out in this thesis and highlights the main contributions. In Section 7.2 explores some future research directions backed up by initial results from pilot works in progress.

7.1 Summary of Main Findings

The investigations reported in this thesis have mainly focused on developing reliable computational tools to deal with two gynaecological problems, first, automatic classification

Ovarian Tumours (Benign or Malignant). Second, is automatic identification of Miscarriage cases from Pregnancy unknown viability PUV in early stage of pregnancy from US images of the Gestational Sac. The aim and objectives of our research work was to complement the automation of certain quantitative parameters that gynaecologist experts and radiologists manually determine using basic computer vision tools, with attributes that encapsulate image-content information that are not directly observable/accessible from the image without applying sophisticated mathematical transformations.

Although ultrasound imaging is considered as one of the most important and effective imaging modality in detecting gynaecological abnormalities, analysing these images is highly dependent on operator experience. Moreover, limitations of the human eye-brain visual system as a result of tiredness and the presence of overlapping and obscured structures in US images may cause detection and/or interpretation errors. An expert observer makes a diagnostic decision with a level of certainty, and if not entirely certain about his/her diagnostic decisions then often other expert opinions are sought and may be essential for diagnosing specific cases. This process will increase the level of confidence and lead to better diagnosis results. Different experts could make different diagnoses which are often labelled as "inconclusive cases" that merit further clinical examinations and perhaps new scans with a different imaging modality. Therefore, there is a need to develop a computer-aided diagnosis system CAD for the purpose of detecting gynaecological abnormality in early stage.

Through regular discussions/consultation with our medical experts in the field, we identified a number of characteristics for the sort of support CAD tool that we should be developing. First of all we need image processing techniques to automate the measurement of certain parameters that experts determine manually and use in their respective diagnosis. Such techniques should improve the efficiency and precision of the measured attributes in order to avoid intra and inter observer variations in manual measurements of the parameters. Motivated by advances in image-based pattern recognition applications we identified additional characteristics for the developed tools to go beyond the confine of automating known parameters by investigating additional measurements as well as alternative and effective discriminating features that can be directly extracted from US images that are "outside" of the known or directly accessible parameters to medical experts. Lastly the accuracy of diagnosis of both gynaecological problems is needed to be supported with a level of confidence. Therefore, we also attempted to make the developed diagnostic tool to be closer to the reality in diagnosis by introducing the confidence level to the diagnosis

outcome. These factors were taken into account when designing the research framework of the investigations conducted, not necessarily in strict linear order, during the life of the project. The work reported in this thesis covered the various components of the framework in the order of presentation and the various experimental works established the validity of the various innovative approaches adopted in successfully meeting the main research objectives.

1. Pre-processing the US images is necessary as a mean to reasonably successful segmentation of ROI. US images often suffer from the effect of speckle noise beside the presence of artefacts that even make ROI manual segmentation a challenge. We used a combination of spatial domain filters and operations as pre-processing procedures to enhance US images for both applications. We have shown that the Non-local means (NLM) filter is effective in reducing speckle noise from US images. We have demonstrates that these procedures help enhance the inner border of malignant tumours and reliably segment the GS as well as the ovarian regions.

2. Successful image pre-processing was only the prelude for the feature extraction step.

    a. We developed reliable automated procedures to extract several types of features to model GS dimensional measurements, few of which are manually determined by radiologist and used by gynaecologists to identify miscarriage cases. Testing the performance of the corresponding automatic diagnostic schemes has undoubtedly contributed to achieving excellent accuracy when the kNN classifier was used. Even the misclassified PUV images turned out to have border line MSD and became cases of miscarriage in follow-up diagnoses, indicating that with some refinement our automatic method can be enabled to give advance prediction.

    b. Recognising the importance of image texture for manual diagnosing of Ovarian Tumours by experts, we introduced and test the performance of several local as well as global image-texture based features in the spatial as well as the frequency domains. The spatial domain features included image histograms, first order statistical features, and versions of LBP. From the frequency domain the binarised FFT spectrum, as a rich source of global image texture information, we extracted the innovative new condensed three features FFGF feature vector that encapsulates the image texture information that depends on all image pixel values. We have demonstrated, with experiments, that each of these spatial/frequency domain image-texture features can define an automatic Ovarian Tumour diagnostic scheme

that have relatively high power of discriminating Benign from Malignant tumours when classified by SVM. The FFGF best performing scheme achieving more than 85% accuracy. These excellent results indicates the potentials of using machine learning + health decision making in the highly specialised branch of biomedical image analysis for gynaecological abnormalities.

3. The influence of different classifiers on the performance of the adopted feature vectors was investigated and it established empirically that for Ovarian Tumour classification, SVM lead to significantly improved performance with the image-texture features (FFGF, LBP, and Histogram) compared to the kNN classifier. For miscarriage identification, however, the GS measurement features (MSD, Perimeter, Volume, etc.) had better accuracy with kNN than with SVM but by a small margin.

4. Introduced a rather simple measure of confidence to quantify the goodness of automatic decision making, primarily for the chosen applications, to emulate real life medical diagnostics from medical radiology images. It differs from the common statistical concept of confidence in that it is calculated with each decision as it is made regardless of decision accuracy rather than an aftermath measure that depend on the statistics of success and failure. Experimental results demonstrated a strong link between this measure and accuracy rates, so that high/medium level of confidence for a decision is most likely accurate while decisions made with low level of confidence could raise alarm. Further work is needed to refine this concept perhaps by introducing more than three levels and/or changing their boundaries.

5. Our investigations of fusion models of multi-feature schemes revealed that feature level fusion yields degraded performance compare to all its single components perhaps due to differences in their nature. But, score level feature fusions (weighted/un-weighted average) have led to improved results and the decision level fusion of three sets of features using majority rule is slightly less successful. For decision level fusion of two sets of features, we attempted to resolve conflicting decisions by using their associated level of confidence measures. This has resulted in cases where it was difficult to resolve the conflict which we labelled as Not Sure (NS) decision. If we consider the NS label as good practice and an incentive to conduct more tests, rather than misclassification, then we get significantly improved accuracy. Again, extensive future research is needed to develop this approach to a mature system that can be part of a wider Decision Support System in this field. Again refining the concept of confidence as discussed above may lead to improved performance.

7.2 Future Work

The developed innovative classification schemes and the proposed fusions strategies and the simple, though useful and medically relevant, concept of level of confidence of a classification decision provide the main ingredient of a computational Decision Support System (DSS) to deal with gynaecological abnormalities with a wider objectives than the ones we dealt with in this thesis. Achieving such a noble goal, however, require expanding the remit of the work reported here and further refined the technological tools and concepts investigated so far. Here we give a brief description of our future work and identify the essential issues that need to be investigated in depth.

The general directions of our future work in both areas will continue to focus on automatic classification of the same two gynaecological abnormalities from US scan images, and build the success we achieved so far in collaboration with gynaecologists toward designing the ultimate CAD that can be deployed in hospitals for a more extensive testing on a significantly larger number of images. We shall now discuss the main components of our future work in the two selected applications.

7.2.1 Ultrasound Images of Ovarian Tumour

Although reasonably high accuracy rate was achieved (in access of 86%) in distinguishing Malignant from Benign tumours, improving accuracy and reducing the NS cases that occur when fusing two features at the decision level is necessary if the scheme is to be embraced by the gynaecology community. Our intended accuracy improving approaches include:

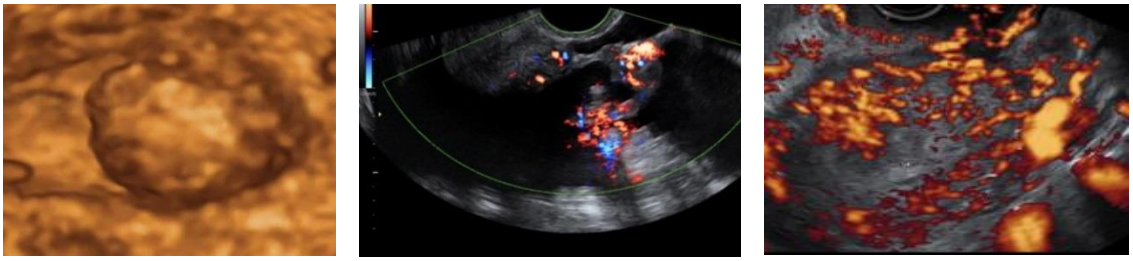
1. Develop new or modified sets of features to model Ovarian Tumour masses using both spatial and frequency image domains. For example we shall extend the extraction of the LBP features into frequency domain with focus on wavelet sub-bands with the possibility of using different versions that rely on shorter patterns. We shall also investigate statistical parameter features obtained from distribution of the high frequency wavelet sub-bands. Another possibility is to use apply the FFGF in blocks rather than in the whole image, or extracting multiple FFGFs ellipses from the binarised FFT spectrum using multiple thresholds.

2. The variety of features mentioned above should provide opportunities to fuse more combinations of features than we have so far with the possibility of using other than the kNN and SVM. Moreover, we should extend the fusion work to include the fusion of multi-classifiers with the same or different types of features.

3. Investigate different ways of reducing the number of NS cases for example by refine the concept of decision conference. Such possibilities include subdividing each of the three levels into sublevels or changing the boundaries between the different levels.

4. Investigate the effectiveness of using different types of US images such as colour Doppler images, and 3D US scans (see Figure 7.1 for examples of such US images).

5. More importantly, extend our work on binary classification of Benign and malignant ovarian tumour, into multi-class Ovarian Tumour identification. In fact, there is a more interest in distinguishing between different types of benign tumours or different stages of malignancy as well as the Borderline Benign/Malignant cases. Note that, the above intended extension of our work in the future will be an essential ingredient for success in this case. In fact, we have conducted some pilot studies using our existing image-texture features for automatic diagnosis of three different types of benign ovarian tumours (Mucinous cystadenoma, Serous cystadenoma, Serous cystadenofibroma). The results these limited investigations are promising and add to our determination to do the above listed set of future work and to extend the application to the various multi-class ovarian tumours identification.



(a) (b) (c)

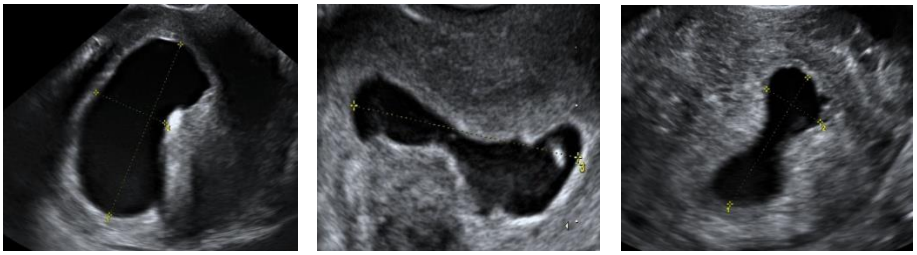
Figure 7.1 Different US images (a) 3D scan of Endometrioma/benign tumour (b) 2D colour Doppler of metatastic malignant, (c) 2D power Doppler for highly vascularized malignant tumour.

7.2.2 Ultrasound Images of Gestational sac

The fact that our experiment on miscarriage classification using few GS dimension measurements should not be used to claim that it works in all cases. Those experiments excluded two important cases where miscarriage may or may not be detected either due to lack of sufficient examples or due to the fact that miscarriage may occur at more advance stage of pregnancy which we did not cover. Hence our future work would focus on those un-investigated cases rather than improve accuracy for the cases covered in this thesis.

1. Expert can predict Miscarriage from the shape of the GS in the first scan. In our investigation, such scans were not sufficiently included in our experimental datasets. If the GS border appears irregular in its shape, then abnormality with GS growth is predicted that

may lead to a miscarriage at a later stage. Figure 7.2 is shows some examples. In order to deal with such cases, other features capturing the characteristics of GS border and shape need to be included in modelling the GS. Sought after features should be able to represents the irregularity of the GS perimeter. We anticipate that perimeter length and significant deviation from circular shape may be useful in dealing with this problem, but first we need to collect sufficiently large number of such samples.



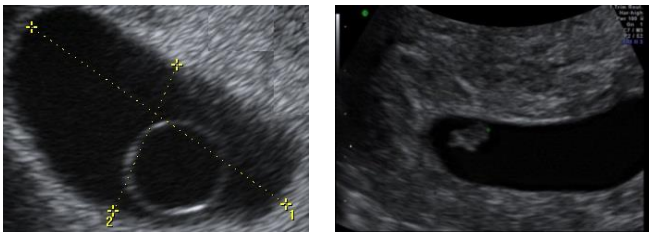
(a) (b) (c)

Figure 7.2 Represents the miscarriage cases that confirm by expert with irregular border

2. Initiate the segmentation, quantification and classification of other stages of miscarriage cases i.e. GS with Yolk or Embryo inside. Figure 7.3 shows example of both these different types of miscarriage in different stages.

3. It would be interesting to investigate the use of 3D ultrasound scan of gestational sac which allows a more accurate estimation of the GS volume and other parameters.

4. Finally, we plan to investigate the effect of external descriptors (demographic) such as age and level of human Chorionic Gonadotropin (hCG) on miscarriage.



(a) (b)

Figure 7.3 Ultrasound images of GS (a) with Yolk Sac (b) with embryo inside

REFERENCES

Acharya, U. R., M. R. K. Mookiah, S. V. Sree, R. Yanti, R. Martis, L. Saba, F. Molinari, S. Guerriero and J. S. Suri (2013). Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification. Ovarian Neoplasm Imaging, Springer: 425-440.

Acharya, U. R., S. Vinitha Sree, L. Saba, F. Molinari, S. Guerriero and J. S. Suri (2012). Ovarian tumor characterization and classification: A class of GyneScan™ systems. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE.

ACOG, T. A. C. o. O. a. G. (2015). "Early Pregnanacy Loss." from http://www.acog.org/-/media/For-Patients/faq090.pdf?dmc=1&ts=20150314T0644120655.

ACS, A. C. S. (2014). "How is ovarian cancer diagnosed?", from http://www.cancer.org/cancer/ovariancancer/detailedguide/ovarian-cancer-diagnosis.

Agrawal R, I. T., Swami A (1993). "Mining association rules between sets of items in large databases." ACM SIGMOD Record 22(2).

Ahonen, T., A. Hadid and M. Pietikäinen (2004). Face recognition with local binary patterns. Computer vision-eccv 2004, Springer: 469-481.

Al-Bayati, M. and A. El-Zaart (2013). "Mammogram Images Thresholding for Breast Cancer Detection Using Different Thresholding Methods."

Alcazar, J. L. (2006). "Tumor angiogenesis assessed by three-dimensional power Doppler ultrasound in early, advanced and metastatic ovarian cancer: a preliminary study." Ultrasound in Obstetrics & Gynecology 28(3): 325-329.

Alcázar, J. L. and M. Jurado (2011). "Three-dimensional ultrasound for assessing women with gynecological cancer: a systematic review." Gynecologic oncology 120(3): 340-346.

Ameye, L., L. Valentin, A. C. Testa, C. Van Holsbeke, E. Domali, S. Van Huffel, I. Vergote, T. Bourne and D. Timmerman (2009). "A scoring system to differentiate malignant from benign masses in specific ultrasound based subgroups of adnexal tumors." Ultrasound in Obstetrics & Gynecology 33(1): 92-101.

Bell, D. A. (1991). "Ovarian surface epithelial-stromal tumors." Human pathology 22(8): 750-762.

Beyer, K., J. Goldstein, R. Ramakrishnan and U. Shaft (1999). When is "nearest neighbor" meaningful? Database Theory—ICDT'99, Springer: 217-235.

Bharati, M. H., J. J. Liu and J. F. MacGregor (2004). "Image texture analysis: methods and comparisons." Chemometrics and intelligent laboratory systems 72(1): 57-71.

Boote, E. J. (2003). "AAPM/RSNA Physics Tutorial for Residents: Topics in US: Doppler US Techniques: Concepts of Blood Flow Detection and Flow Dynamics 1." Radiographics 23(5): 1315-1327.

Bottomley, C., V. Van Belle, E. Kirk, S. Van Huffel, D. Timmerman and T. Bourne (2013). "Accurate prediction of pregnancy viability by means of a simple scoring system." Human Reproduction 28(1): 68-76.

Bourne, T. and C. Bottomley (2012). "When is a pregnancy nonviable and what criteria should be used to define miscarriage?" Fertility and sterility 98(5): 1091-1096.

Braem, M. G., N. C. Onland-Moret, L. J. Schouten, R. F. Kruitwagen, A. Lukanova, N. E. Allen, P. A. Wark, A. Tjønneland, L. Hansen and C. M. Braüner (2012). "Multiple miscarriages are associated with the risk of ovarian cancer: results from the European Prospective Investigation into Cancer and Nutrition." PloS one 7(5): e37141.

Brown, D., D. Emerson, R. Felker, M. Cartier and W. Smith (1990). "Diagnosis of early embryonic demise by endovaginal sonography." Journal of ultrasound in medicine 9(11): 631-636.

Brown,H.L.(2014)."Stage of Development of the Fetus." from http://www.merckmanuals.com/home/women-s-health-issues/normal-pregnancy/stages-of-development-of-the-fetus.

Buades, A., B. Coll and J.-M. Morel (2005). A non-local algorithm for image denoising. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE.

Buades, A., B. Coll and J. M. Morel (2004). "On image denoising methods." CMLA Preprint 5.

Bushberg, J. T. and J. M. Boone (2011). The essential physics of medical imaging, Lippincott Williams & Wilkins.

Campion, E. W., P. M. Doubilet, C. B. Benson, T. Bourne and M. Blaivas (2013). "Diagnostic criteria for nonviable pregnancy early in the first trimester." New England Journal of Medicine 369(15): 1443-1451.

Castellano, G., L. Bonilha, L. Li and F. Cendes (2004). "Texture analysis of medical images." Clinical radiology 59(12): 1061-1069.

Chakkarwar, V., M. S. Joshi and P. S. Revankar (2010). Automated analysis of gestational sac in medical image processing. Advance Computing Conference (IACC), 2010 IEEE 2nd International, IEEE.

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

Chan, V. and A. Perlas (2011). Basics of ultrasound imaging. Atlas of Ultrasound-Guided Procedures in Interventional Pain Management, Springer: 13-19.

Chen, V. W., B. Ruiz, J. L. Killeen, T. R. Coté, X. C. Wu, C. N. Correa and H. L. Howe (2003). "Pathology and classification of ovarian tumors." Cancer 97(S10): 2631-2642.
Cortes, C. and V. Vapnik (1995). "Support-vector networks." Machine learning 20(3): 273-297.

Cristianini, N. and J. Shawe-Taylor (2000). An introduction to support vector machines and other kernel-based learning methods, Cambridge university press.

Czerwinski, R. N., D. L. Jones and W. D. O'Brien Jr (1999). "Detection of lines and boundaries in speckle images-application to medical ultrasound." Medical Imaging, IEEE Transactions on 18(2): 126-136.

Dai, S. Y., K. Hata, E. Inubashiri, K. Kanenishi, A. Shiota, M. Ohno, Y. Yamamoto, Y. Nishiyama, M. Ohkawa and T. Hata (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): 364-370.

Dangeti, S. (2003). Denoising techniques-a comparison, Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering in The Department of Electrical and Computer Engineering by Sarita Dangeti BE, Andhra University College of Engineering.

Dhawan, A. P. (2011). Medical image analysis, John Wiley & Sons.

Duda, R. O., P. E. Hart and D. G. Stork (2001). "Pattern Classification. JohnWiley & Sons." New York.

Edler, I. and K. Lindström (2004). "The history of echocardiography." Ultrasound in medicine & biology 30(12): 1565-1644.

Edwards, R. and P. Steptoe (1975). "Induction of follicular growth, ovulation and luteinization in the human ovary." Journal of reproduction and fertility. Supplement(22): 121-163.

England, H. a. S. C. I. C. N. m. s. (London HSCIC, 2013). "NHS maternity statistics: England 2011-2012",from:http://www.hscic.gov.uk/article/2021/Website-Search?productid=10061&q=miscarriage&sort=Relevance&size=10&page=3&area=both.

Fenster, A. and D. B. Downey (1996). "3-D ultrasound imaging: A review." Engineering in Medicine and Biology Magazine, IEEE 15(6): 41-51.

Findlay, J., J. Kerr, K. Britt, S. Liew, E. Simpson, D. Rosairo and A. Drummond (2009). "Ovarian physiology: follicle development, oocyte and hormone relationships." Anim. Reprod 6(1): 16-19.

Fishman, D. A., L. Cohen, S. V. Blank, L. Shulman, D. Singh, K. Bozorgi, R. Tamura, I. Timor-Tritsch and P. E. Schwartz (2005). "The role of ultrasound evaluation in the detection of early-stage epithelial ovarian cancer." American journal of obstetrics and gynecology 192(4): 1214-1221.

Fix, E. and J. L. Hodges Jr (1951). Discriminatory analysis-nonparametric discrimination: consistency properties, DTIC Document.

Garra, B. S., B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun and R. K. Zeman (1993). "Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis." Ultrasonic Imaging 15(4): 267-285.

GE,h.(2014)."UltrasoundTransducers."from http://www3.gehealthcare.com/en/products/categories/ultrasound/ultrasound_probes.
Geirsson, R. and R. Busby-Earle (1991). "Certain dates may not provide a reliable estimate of gestational age." BJOG: An International Journal of Obstetrics & Gynaecology 98(1): 108-109.

Giger, M. L., H.-P. Chan and J. Boone (2008). "Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM." Medical physics 35(12): 5799-5820.

Gonzalez, R. C., R. E. Woods and S. L. Eddins (2004). Digital image processing using MATLAB, Pearson Education India.

Gramellini, D., S. Fieni, L. Sanapo, G. Casilla, C. Verrotti and G. B. Nardelli (2008). "Diagnostic accuracy of IOTA ultrasound morphology in the hands of less experienced sonographers." Australian and New Zealand Journal of Obstetrics and Gynaecology 48(2): 195-201.

Guyon, I. and A. Elisseeff (2006). An introduction to feature extraction. Feature extraction, Springer: 1-25.

Hamid, B. A. (2011). Image texture analysis of transvaginal ultrasound in monitoring ovarian cancer, School of Engineering, Cardiff University.

Hamlett, W. C. and T. J. Koob (1999). "Female reproductive system." Sharks, skates, and rays: the biology of elasmobranch fishes: 398-443.

Han, J. and M. Kamber (2001). Data mining: concepts and techniques, Morgan Kaufmann San Francisco, Calif, USA.

Hangiandreou, N. J. (2003). "AAPM/RSNA Physics Tutorial for Residents: Topics in US: B-mode US: Basic Concepts and New Technology 1." Radiographics 23(4): 1019-1033.
Haralock, R. M. and L. G. Shapiro (1991). Computer and robot vision, Addison-Wesley Longman Publishing Co., Inc.

Harrington, N. (2007). "Segmentation of human ovarian follicles from ultrasound images acquired in vivo using geometric active contour models and a naïve Bayes classifier."

Hiransakolwong, N., K. Hua, K. Vu and P. S. Windyga (2003). Segmentation of ultrasound liver images: an automatic approach. Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, IEEE.

Hiremath, P. and J. R. Tegnoor (2010). Automatic detection of follicles in ultrasound images of ovaries using active contours method. Proceedings of IEEE international conference on computational intelligence and computing research (ICCIC-2010).

Hiremath, P. and J. R. Tegnoor (2010). "Automatic detection of follicles in ultrasound images of ovaries using edge based method." IJCA, Special Issue on RTIPPR 2: 120-125.

Hiremath, P. and J. R. Tegnoor (2013). "Follicle Detection and Ovarian Classification in Digital Ultrasound Images of Ovaries." Advancements and Breakthroughs in Ultrasound Imaging, InTechOpen, UK: 167-199.

Hoskins, P. R., K. Martin and A. Thrush (2010). Diagnostic ultrasound: physics and equipment, Cambridge University Press.

Huang, Y. L., D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu and W. Moon (2008). "Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound." Ultrasound in Obstetrics & Gynecology 32(4): 565-572.

Jeong, Y.-Y., E. K. Outwater and H. K. Kang (2000). "Imaging Evaluation of Ovarian Masses 1." Radiographics 20(5): 1445-1470.

Jokubkiene, L., P. Sladkevicius and L. Valentin (2007). "Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses?" Ultrasound in Obstetrics & Gynecology 29(2): 215-225.

Jurkovic, D., C. Overton and R. Bender-Atik (2013). "Diagnosis and management of first trimester miscarriage." BMJ 346: f3676.

Kaijser, J., T. Bourne, L. Valentin, A. Sayasneh, C. Van Holsbeke, I. Vergote, A. C. Testa, D. Franchi, B. Van Calster and D. Timmerman (2013). "Improving strategies for diagnosing ovarian

cancer: a summary of the International Ovarian Tumor Analysis (IOTA) studies." Ultrasound in Obstetrics & Gynecology 41(1): 9-20.

Kaur, A. and A. Kaur (2011). "Transvaginal ultrasonography in first trimester of pregnancy and its comparison with transabdominal ultrasonography." Journal of Pharmacy And Bioallied Sciences 3(3): 329.

Kervrann, C. a. H., F (1995). "A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics." Image Processing, IEEE Transactions 4(6): pp.856-862.

Kinkel, K., H. Hricak, Y. Lu, K. Tsuda and R. A. Filly (2000). "US Characterization of Ovarian Masses: A Meta-Analysis 1." Radiology 217(3): 803-811.

Krivanek, A. and M. Sonka (1998). "Ovarian ultrasound image analysis: Follicle segmentation." Medical Imaging, IEEE Transactions on 17(6): 935-944.

Kurani, A. S., D.-H. Xu, J. Furst and D. S. Raicu (2004). Co-occurrence matrices for volumetric data. 7th IASTED International Conference on Computer Graphics and Imaging, Kauai, USA.

Kurjak, A. and S. Kupesic (2003). Color Doppler and 3D Ultrasound in Gynecology, Infertility and Obstetrics, JAYPEE BROTHERS PUBLISHERS.

Kurtz, A. B., J. V. Tsimikas, C. M. Tempany, U. M. Hamper, P. H. Arger, R. L. Bree, R. J. Wechsler, I. R. Francis, J. E. Kuhlman and E. S. Siegelman (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 1." Radiology 212(1): 19-27.

Lee, C. and D. Landgrebe (1993). "Feature extraction and classification algorithms for high dimensional data." ECE Technical Reports: 212.

Lefebvre, F., M. Meunier, F. Thibault, P. Laugier and G. Berger (2000). "Computerized ultrasound B-scan characterization of breast nodules." Ultrasound in medicine & biology 26(9): 1421-1428.

Levi, C., E. Lyons and D. Lindsay (1990). "Ultrasound in the first trimester of pregnancy." Radiologic clinics of North America 28(1): 19-38.

Levine, D., D. L. Brown, R. F. Andreotti, B. Benacerraf, C. B. Benson, W. R. Brewster, B. Coleman, P. DePriest, P. M. Doubilet and S. R. Goldstein (2010). "Management of asymptomatic ovarian and other adnexal cysts imaged at US: Society of Radiologists in Ultrasound Consensus Conference Statement 1." Radiology 256(3): 943-954.

Levine, M. D. (1985). Vision in man and machine, McGraw-Hill New York.

Liao, R., T. Wan and Z. Qin (2011). Classification of benign and malignant breast tumors in ultrasound images based on multiple sonographic and textural features. Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on, IEEE.

Loizou, C. P. and C. S. Pattichis (2008). "Despeckle filtering algorithms and software for ultrasound imaging." Synthesis lectures on algorithms and software in engineering 1(1): 1-166.

Loue, S. and M. Sajatovic (2004). Encyclopedia of women's health, Springer Science & Business Media.

Luo, L.-Y., D. Katsaros, A. Scorilas, S. Fracchioli, R. Bellino, M. van Gramberen, H. de Bruijn, A. Henrik, U.-H. Stenman and M. Massobrio (2003). "The serum concentration of human kallikrein 10 represents a novel biomarker for ovarian cancer diagnosis and prognosis." Cancer research 63(4): 807-811.

Mäenpää, T. and M. Pietikäinen (2005). "Texture analysis with local binary patterns." Handbook of Pattern Recognition and Computer Vision 3: 197-216.

Malathi, G. and V. Shanthi (2010). Histogram based classification of ultrasound images of placenta. IJCA, Citeseer.

Manay, S. and A. Yezzi (2003). "Anti-geometric diffusion for adaptive thresholding and fast segmentation." Image Processing, IEEE Transactions on 12(11): 1310-1323.

Mason, R. J., V. C. Broaddus, T. Martin, T. E. King Jr, D. Schraufnagel, J. F. Murray and J. A. Nadel (2010). Murray and Nadel's textbook of respiratory medicine: 2-volume set, Elsevier Health Sciences.

Materka, A. (2004). "Texture analysis methodologies for magnetic resonance imaging." Dialogues in clinical neuroscience 6(2): 243.

Materka, A. and M. Strzelecki (1998). "Texture analysis methods–a review." Technical university of lodz, institute of electronics, COST B11 report, Brussels: 9-11.

Mathias, J., P. Tofts and N. Losseff (1999). "Texture analysis of spinal cord pathology in multiple sclerosis." Magnetic Resonance in Medicine 42(5): 929-935.

Michailovich, O. V. and A. Tannenbaum (2006). "Despeckling of medical ultrasound images." Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on 53(1): 64-78.

Mojsilovic, A., M. Popović, S. Marković and M. Krstić (1998). "Characterization of visually similar diffuse diseases from B-scan liver images using nonseparable wavelet transform." Medical Imaging, IEEE Transactions on 17(4): 541-549.

Moore, R. G., A. K. Brown, M. C. Miller, S. Skates, W. J. Allard, T. Verch, M. Steinhoff, G. Messerlian, P. DiSilvestro and C. Granai (2008). "The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass." Gynecologic oncology 108(2): 402-408.

Moradi, M., P. Mousavi, A. H. Boag, E. E. Sauerbrei, D. R. Siemens and P. Abolmaesumi (2009). "Augmenting detection of prostate cancer in transrectal ultrasound images using SVM and RF time series." Biomedical Engineering, IEEE Transactions on 56(9): 2214-2224.

Morris, D. (1988). "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): 387-395.

Myers, E. R., L. A. Bastian, L. J. Havrilesky, S. L. Kulasingam, M. S. Terplan, K. E. Cline, R. N. Gray and D. C. McCrory (2006). "Management of adnexal mass." Evidence report/technology assessment(130): 1-145.

Nailon, W. H. (2010). Texture analysis methods for medical image characterisation, INTECH Open Access Publisher.

NHS, N. H. S. (2015). "A miscarriage is the loss of a pregnancy during the first 23 weeks." from http://www.nhs.uk/Conditions/Miscarriage/Pages/Introduction.aspx.

Nixon, M. (2008). Feature extraction & image processing, Academic Press.

Ojala, T., M. Pietikäinen and D. Harwood (1996). "A comparative study of texture measures with classification based on featured distributions." Pattern recognition 29(1): 51-59.

Ojala, T., M. Pietikäinen and T. Mäenpää (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(7): 971-987.

Omaha Nebraska, C. u. m. c. C. (2013). "Ultrasound of Early Pregnancy." from http://web.archive.org/web/20070814054851/http://radiology.creighton.edu/pregnancy.htm#section4.

Otsu, N. (1975). "A threshold selection method from gray-level histograms." Automatica 11(285-296): 23-27.

Özgen, C. (2011). A medical image processing and analysis framework, middle east technical university.

Palmer, P. E. (1995). Manual of diagnostic ultrasound, World Health Organization.
Pang-Ning Tan, M. S., Vipin Kumar (2005). Introduction to Data Mining, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

Pathak, S. D., D. Haynor and Y. Kim (2000). "Edge-guided boundary delineation in prostate ultrasound images." Medical Imaging, IEEE Transactions on 19(12): 1211-1219.

Permuth-Wey, J. and T. A. Sellers (2009). Epidemiology of ovarian cancer. Cancer Epidemiology, Springer: 413-437.

Pexsters, A., A. Daemen, C. Bottomley, D. Van Schoubroeck, L. De Catte, B. De Moor, T. D'Hooghe, C. Lees, D. Timmerman and T. Bourne (2010). "New crown–rump length curve based on over 3500 pregnancies." Ultrasound in Obstetrics & Gynecology 35(6): 650-655.

Pietikäinen, M., A. Hadid, G. Zhao and T. Ahonen (2011). Local binary patterns for still images. Computer Vision Using Local Binary Patterns, Springer: 13-47.

Rafiee, A., A. Salimi and A. R. Roosta (2008). "A novel prostate segmentation algorithm in TRUS images." World Academy of Science, Engineering and Technology 45: 120-124.

RCOG, R. C. o. O. a. G. (2003). "Ovarian Cysts in Postmenopausal Women." from https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg34/.

RCOG, R. C. o. O. a. G. (2015). "Early miscarriage." from https://www.rcog.org.uk/globalassets/documents/patients/patient-information-leaflets/pregnancy/early-miscarriage.pdf.

RL, V. W., J. Bogumil, I. Dyrenfurth, M. Ferin, R. Jewelewicz, M. Warren, R. Rizkallah and G. Mikhail (1971). "Mechanisms regulating the menstrual cycle in women." Recent Progress in Hormonal Research 26: 63-103.

Rocha, R., A. Campilho, J. Silva, E. Azevedo and R. Santos (2011). "Segmentation of ultrasound images of the carotid using RANSAC and cubic splines." Computer methods and programs in biomedicine 101(1): 94-106.

Rose, R. J. and S. Allwin (2013). "Computerized cancer detection and classification using ultrasound images: a survey." Int. J. Eng. Res. Dev 5: 36-47.

Sayasneh, A., L. Wynants, J. Preisler, J. Kaijser, S. Johnson, C. Stalder, R. Husicka, Y. Abdallah, F. Raslan and A. Drought (2013). "Multicentre external validation of IOTA prediction models and RMI by operators with varied training." British journal of cancer 108(12): 2448-2454.

Setarehdan, S. K. and S. Singh (2012). Advanced algorithmic approaches to medical image segmentation: state-of-the-art applications in cardiology, neurology, mammography and pathology, Springer Science & Business Media.

Sheppard, M. and L. Shih (2005). Efficient image texture analysis and classification for prostate ultrasound diagnosis. Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts. IEEE, IEEE.

Shung, K. K. (2015). Diagnostic ultrasound: Imaging and blood flow measurements, CRC press.

Smutek, D., R. Šára, P. Sucharda, T. Tjahjadi and M. Švec (2003). "Image texture analysis of sonograms in chronic inflammations of thyroid gland." Ultrasound in medicine & biology 29(11): 1531-1543.

Smythe, D. (2004). 3D/4D ultrasound imaging system, Google Patents.

Sohail, A. S. M., M. M. Rahman, P. Bhattacharya, S. Krishnamurthy and S. P. Mudur (2010). Retrieval and classification of ultrasound images of ovarian cysts combining texture features and histogram moments. Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on, IEEE.

Stany, M. P., G. L. Maxwell and G. S. Rose (2010). "Clinical decision making using ovarian cancer risk assessment." American Journal of Roentgenology 194(2): 337-342.

Suetens, P. (2009). Fundamentals of medical imaging, Cambridge university press.
Tan, D. S., R. Agarwal and S. B. Kaye (2006). "Mechanisms of transcoelomic metastasis in ovarian cancer." The lancet oncology 7(11): 925-934.

Tatjana-Mihaela. (2002). "POLYCYSTIC OVARIAN SYNDROME, ENDOMETRIOSIS & SUCCESFUL NATURAL CURE." from http://tatjana mihaela.hubpages.com/hub/polycysticOS#.

Testa, A., S. Ajossa, G. Ferrandina, E. Fruscella, M. Ludovisi, M. Malaggese, G. Scambia, G. Melis and S. Guerriero (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 & Gynecology 26(1): 67-72.

Thaipanich, T. and C.-C. J. Kuo (2010). An adaptive nonlocal means scheme for medical image denoising. SPIE Medical Imaging, International Society for Optics and Photonics.

Timmerman, D. (2000). "Lack of standardization in gynecological ultrasonography." Ultrasound in Obstetrics & Gynecology 16(5): 395-398.

Timmerman, D., P. Schwärzler, W. Collins, F. Claerhout, M. Coenen, F. Amant, I. Vergote and T. Bourne (1999). "Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience." Ultrasound in Obstetrics & Gynecology 13(1): 11-16.

Timmerman, D., A. C. Testa, T. Bourne, L. Ameye, D. Jurkovic, C. Van Holsbeke, D. Paladini, B. Van Calster, I. Vergote and S. Van Huffel (2008). "Simple ultrasound-based rules for the diagnosis of ovarian cancer." Ultrasound in Obstetrics & Gynecology 31(6): 681-690.

Timmerman, D., A. C. Testa, T. Bourne, E. Ferrazzi, L. Ameye, M. L. Konstantinovic, B. Van Calster, W. P. Collins, I. Vergote and S. Van Huffel (2005). "Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group." Journal of Clinical Oncology 23(34): 8794-8801.

Timmerman, D., B. Van Calster, A. C. Testa, S. Guerriero, D. Fischerova, A. Lissoni, C. Van Holsbeke, R. Fruscio, A. Czekierdowski and D. Jurkovic (2010). "Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group." Ultrasound in Obstetrics & Gynecology 36(2): 226-234.

Togashi, K. (2003). "Ovarian cancer: the clinical role of US, CT, and MRI." European radiology 13(6): L87-L104.

Tsai, D.-Y. and K. Kojima (2005). "Measurements of texture features of medical images and its application to computer-aided diagnosis in cardiomyopathy." Measurement 37(3): 284-292.

Twickler, D. M. and E. Moschos (2010). "Ultrasound and assessment of ovarian cancer risk." American Journal of Roentgenology 194(2): 322-329.

UK, C. R. (2011). "Cancer Stats Key Facts Ovarian Cancer." from http://www.cancerresearchuk.org/cancerinfo/prod_consump/groups/cr_common/@nre/@sta/documents/generalcontent/crukmig_1000ast-3058.pdf.

UK, C. R. (2014). "Ovarian cancer incidence statistics." from http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#undefined.

UK, C. R. (2014). "Types of ovarian cancer." from http://www.cancerresearchuk.org/about-cancer/type/ovarian-cancer/about/types-of-ovarian-cancer.

UK, C. R. (2015). "Ovarian cancer." from http://www.cancerresearchuk.org/about-cancer/type/ovarian-cancer/.

Vaes, E., R. Manchanda, P. Autier, R. Nir, D. Nir, H. Bleiberg, A. Robert and U. Menon (2012). "Differential diagnosis of adnexal masses: sequential use of the risk of malignancy index and HistoScanning, a novel computer-aided diagnostic tool." Ultrasound in Obstetrics & Gynecology 39(1): 91-98.

Van Calster, B., D. Timmerman, T. Bourne, A. C. Testa, C. Van Holsbeke, E. Domali, D. Jurkovic, P. Neven, S. Van Huffel and L. Valentin (2007). "Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125." Journal of the National Cancer Institute 99(22): 1706-1714.

Van Calster, B., K. Van Hoorde, W. Froyman, J. Kaijser, L. Wynants, C. Landolfo, C. Anthoulakis, I. Vergote, T. Bourne and D. Timmerman (2015). "Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors." Facts, views & vision in ObGyn 7(1): 32.

Wang, Y., K. Itoh, N. Taniguchi, H. Toei, F. Kawai, M. Nakamura, K. Omoto, K. Yokota and T. Ono (2002). "Studies on tissue characterization by texture analysis with co-occurrence matrix method using ultrasonography and CT imaging." Journal of Medical Ultrasonics 29(4): 211-223.

Webb, A. R. (2003). Statistical pattern recognition, John Wiley & Sons.

WiseGeek. (2013). "Medical Imaging." from http://www.wisegeekhealth.com/what-is-medical-imaging.htm.

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(10): 6737-6741.

Xie, J., Y. Jiang, H.-T. Tsui and P.-A. Heng (2006). "Boundary enhancement and speckle reduction for ultrasound images via salient structure extraction." Biomedical Engineering, IEEE Transactions on 53(11): 2300-2309.

Yang, J., J. Fan, D. Ai, S. Zhou, S. Tang and Y. Wang (2015). "Brain MR image denoising for Rician noise using pre-smooth non-local means filter." Biomedical engineering online 14(1): 2.

Zhang, L., S. Chen, S. Li and T. Wang (2011). Automatic measurement of early gestational sac diameters from one scan session. SPIE Medical Imaging, International Society for Optics and Photonics.

Zhang, X., G. Hou, J. Ma, W. Yang, B. Lin, Y. Xu, W. Chen and Y. Feng (2014). "Denoising MR images using non-local means filter with combined patch and pixel similarity."

Zhang, Y.-J. (2006). Advances in image and video segmentation, IGI Global.

Zhu, C., J. Ni, Y. Li and G. Gu (2009). Speckle noise suppression techniques for ultrasound images. Internet Computing for Science and Engineering (ICICSE), 2009 Fourth International Conference on, IEEE.