Texture Analysis based Machine Learning Algorithms For Ultrasound Ovarian Tumour Image Classification within Clinical Practices

By

Dhurgham Al-karawi

School of Computing

University of Buckingham

United Kingdom

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

September 2019

Abstract

Research investigations reported in this thesis, aim to contribute to the efforts of developing reliable ovarian tumour classification tools using texture features extracted from B-mode ultrasound ovarian tumour scan images. This kind of research is necessitated by the shortage of highly trained sonographers and gynaecologists in order to reduce the heavy pressure on healthcare systems throughout the world. Our ultimate aim is to automate the error-prone process of the laborious manual examination of the ultrasound scan images, and we, therefore, exploit advances in Machine learning and computer vision to develop informative software to be integrated within clinical setup.

Our research was guided by an extensive literature review of existing research in this and related fields, building on existing collaborations with medical expertise, and evidence from systems biology research that carcinogenesis results in changing the texture of cysts cellular network. These considerations led to adapting image texture analysis approaches as an adequate source for Machine learning algorithms and software tools. Most existing research works in general biomedical image-based diagnostics are directed towards identifying one or few best performing texture features. Instead, our analysis aimed at extracting a suit of texture-based image features that together contribute to effective ultrasound ovarian tumour image classification. This open-minded strategy unearthed a plethora of texture-based features and in different image domains beyond the spatial domain, which depicts a visual image of the scanned tissue. There is a significant variation in the dimensionality of the texture features, included in our investigations, and although we use different well-known classifiers in evaluating performances, the focus of the comparisons made are not on the choice of classifiers.

This thesis includes many contributions; the most significant ones can be summarised as follows:

1. Established that even without pre-processing the scanned images spatial domain is a rich source of 7 microscale texture primitives that can distinguish malignant tumour scans from benign ones with accuracy well above being a case of random chance prediction (70% -83%). The simple majority rule fusion of an odd number of features yield accuracy in the range 83% - 90%.
2. Developed a smart adaptive speckle-noise reduction scheme that applies noise reduction in blocks of the cropped tumour images (not the entire image) only if

(Skewness, Kurtosis) pair in the block satisfies a criterion determined by training. This adaptive pre-processing is shown to significantly improve the performance of all investigated texture schemes, not only the spatial domain ones.

3. Modified the existing frequency domain texture feature (FFGF), by adaptively pre-processing the cropped tumour image prior to computing its Fourier Spectrum, and using a different binarization scheme to extract the bright elliptical shape at the centre of the FFT spectrum. These modifications improved the accuracy of the original FFGF scheme 85.9% to more than 92%.

4. When attempted to reduce the dependencies between the 3 ellipse parameters of FFGF has shown that even better accuracy (> 95%) can be achieved using a single parameter (the minor axes). These results led to establishing that the FFT-spectrum image is a very rich source of texture information only obscured by its somewhat visually "meaningless" display. We found that all of the features extracted from the FFT-spectrum outperform their spatial domain counterpart, and the fusion of the 7 FFT-spectrum based schemes achieved accuracy of > 97.5%.

5. We further extended the list of texture-based image features beyond the spatial domain and beyond the FFGF schemes by extracting some of the previously defined texture features not only from the FFT-spectrum but also in any image transform domain such as the LBP domain. Again, the texture features in the LBP domain outperformed the spatial domain counterparts, with FFGF from the LBP domain achieving accuracy of 94% which even outperforms the modified FFGF.

6. Finally, the extensive experiments simply opened a Pandora Box of image textures. Instead, of continue other image transform domain, we created two versions of an ML-based software that incorporate 9 spatial domain texture-based features (the original 7 + the Skewness + the Kurtosis) to be used for a prospective test of 100 fresh cases, collected and examined histologically by an IOTA expert gynaecologist at Queen Charlotte and Hammersmith Hospital in London during the period (Oct 2018 – Jan 2019). Version 2 incorporates the smart adaptive speckle noise removal resulted in an accuracy of 94%.

Acknowledgement

I am indebted to my supervisors, **Professor Sabah A. Jassim** and **Dr Hisham Al-Assam**. Without their support and encouragement, this thesis and the research it reports would not have been possible. **Professor Jassim,** in particular, provided much-needed guidance regarding computational and mathematical subjects and tirelessly provided suggestions from the start of my research journey to the end of the project. **Dr Al-Assam** supported me throughout my studies and provided me with meaningful ideas and encouragement. They made me feel confident and able to overcome the difficulties encountered.

Special thanks go to **Mrs Sharon Salerno** and **Mrs Jayne Kelly**, Administrator of Psychology and Applied Computing Departments for her support and assistance. Moreover, Special thanks go to **Mr Hongbo Du** for his support and help, School of Computing. Also, thanks to **Professor Tome Bourne, Professor Drik Timmerman, Dr Ahmed Sayasneh** and **Dr Chiara Landolfo** for providing the dataset and guiding with medical information.

Regarding my family, I am deeply indebted to my parents. They supported me financially and emotionally and my mother ceaselessly kept me in her thoughts and prayers as I studied far from home. My brothers and sisters also listened to me when I was troubled and encouraged me. My sister **Dr Susan** shared her experiences completing her PhD and made me feel like there was someone who understood what I was going through. Her husband, **Dr Marcus**, also provided me with calm advice and reminded me that everyone struggles to do their PhD but that in time the goal will be achieved and to be patient and continue doing my best.

I have to thank all my friends who have supported and encouraged me during my study. Special thanks go to **Dr Ali Jaziny** and his family. I will never forget his support and will always remember our first meeting. Furthermore, I would like to thanks **Mr Aras Assed** and **Mr Mohammed Shwany** and their families for personal support and lovely times we all have had together. Also, Special thanks go to **Dr Dheyaa Ibrahim** for his support.

Finally, Special thanks for the financial support from **the Government of Iraq / Ministry of Higher Education and Scientific Research Iraq**, who granted my study.

Dedication

This Work Is Humbly Dedicated To My Wonderful

Parents, Brothers and Sister.

Dhurgham Al-karawi

Declaration

I hereby declare that this thesis, submitted to the University of Buckingham as a fulfilment of requirements for the degree of Doctor of Philosophy in Computer Science has not been submitted as an exercise for a similar degree at any other university. Furthermore, I certify that the work described in this thesis is completely my own except for excerpts and summaries, whose sources are appropriately cited in the references.

This thesis might be made available within the university library and perhaps photocopied or loaned to other libraries for the purposes of consultation.

Dhurgham Al-karawi

Table of Contents

Abstract ..... **i**
Acknowledgement ..... **iii**
Dedication ..... **iv**
Declaration ..... **v**
List of Figures ..... **x**
List of Tables ..... **xiii**
Abbreviations ..... **xiv**
Chapter 1 ..... **1**
Introduction **1**
1.1 Overview 1
1.2 Research Motivation 3
1.3 Research Aim and Objectives 4
1.4 Research Framework of the Proposed Ovarian Tumour Classification 6
1.5 Thesis Contributions 7
1.6 List of Publications 8
1.7 Dissertation Layout 8
Chapter 2 ..... **10**
Medical and Computational Background **10**
2.1 Medical Imaging Systems 10
2.1.1 Medical Ultrasound Image 11
2.1.2 The Equipment of Ultrasound 12
2.1.3 Scanning Process 12
2.2 Background of Human Female Reproductive System 14
2.2.1 Overview of Ovarian Cancer 15
2.2.2 Histological Types of Ovarian Tumours 16
2.2.3 Stages of Ovarian Cancer 17
2.2.4 Diagnosing Ovarian Tumours - Overview 18
2.2.4.1 Role of Ultrasound Imaging in Ovarian Cancer Diagnosis 19
2.3 Design and Development of Automatic Medical Imaging-based Diagnostic Systems 20
2.3.1 Image Pre-processing 21
2.3.1.2 Speckle Noise and its Pattern and Physical Properties 22
2.3.2 Segmentation of US Images of Ovarian Tumour 23

2.3.3 Feature Representation of Images 23
2.3.3.1 Texture-Based Image Features and Analysis 24
2.4 Classification Techniques 26
2.4.1 Support Vector Machine (SVM) 26
2.4.2 k-Nearest Neighbour (kNN) Classifier 28
2.5 Summary 29
Chapter 3 31
Computerised Ovarian Tumor Diagnostic – A Literature Review 31
3.1. Ultrasound image Pre-processing 31
3.1.1 Image Noise - Nature characteristics 32
3.2 A Review of Image De-Noising Algorithms 33
3.3 Segmentation of US Images of Ovarian Tumour 35
3.4 Texture Representation of Segmented Ultrasound Images 40
3.4.1 Statistical based Image Texture descriptors 41
3.4.2 Transform-based Texture Descriptors 42
3.4.3 Geometric based Texture Descriptors 43
3.5 Existing Computer-based Ultrasound Ovarian Tumour Diagnosis 44
3.7 Summary 48
Chapter 4 49
Texture-based Ultrasound image Analysis for Ovarian Tumor classification 49
4.1 Introduction 49
4.2 Ultrasound Image Texture Feature Descriptors 50
4.2.1. Statistics Histogram Properties 50
4.2.2 Local Binary Pattern LBP 52
4.2.3 Gabor Filter 54
4.2.4 Histograms of Oriented Gradient 54
4.2.5 Fractal Dimension (FD) 55
4.2.6 Hu's Invariant Moments (IM) 57
4.3. Performance Evaluation of Texture-based Schemes 58
4.3.1. The Experimental Dataset 59
4.3.2. Evaluation Protocols 60
4.3.3. Performance Measures 60
4.4. Experimental Results and Discussion 61
4.4.1. Performance of Single Texture-Feature based Scheme 61
4.4.2 Decision-based Fusion of Multiple Schemes 64

4.5 Summary 66
Chapter 5 67
Mitigating Factors Influencing the Performance of the Texture-based Ovarian Tmours Diagnostic Schemes 67
5.1 Effect of Speckle-Noise on the Classification of US Ovary Scans 67
5.1.1 Speckle Noise in Ultrasound Images 68
5.1.2 Impact of Speckle Noise in Ultrasound Images on Diagnosis 69
5.2 Trainable Systems for Adaptive Speckle Noise Removal 72
5.2.1 Model 1: Adaptive Global Speckle Noise Reduction 73
5.2.2 Model 2: Localised and Adaptive Speckle Noise Reduction 75
5.3 Adaptive Pre-processing – Experiment and Discussion 76
5.4 The Impact of RoI Size Variation on the Performance of Texture-based Schemes 80
5.5 A Proposed Size Variation Mitigating Method 82
5.6 Summary 84
Chapter 6 86
Classification in Frequency and Transform Domains 86
6.1 Fourier Transform-based Image Texture Features 87
6.1.1 Fourier Spectrum of Ultrasound Images 87
6.1.2 Geometric FFT based Diagnostic Schemes 90
6.2 Experimental Results and Discussion 94
6.3 Texture Analysis of FFT Spectrum Using Other Features 97
6.4 Texture-based Diagnostic Schemes in the LBP Transform Domains 101
6.5 Summary 104
Chapter 7 106
A Prospective Clinical Test of the developed schemes 106
7.1 Ovarian Tumor texture-based Software for Clinical Testing 107
7.1.1 Model Selection for the Prospective Test 107
7.1.2 The Clinical Test Methodology 108
7.1.3 The Software Interface 108
7.2 Experimental Work and Prospective Test Results 110
7.2.1 Performance of Version 1 110
7.2.2 Performance of Version 2 111
7.3 Deep Learning CNN System for Ovarian Tumour Classification -A Pilot Study 114

7.3.1 Deep Learning 115
7.3.1.1 Convolutional Neural Networks 115
7.3.2 Training of CNN's 117
7.4 The Adopted CNN Architecture in This Pilot Study 119
7.4.1 Experimental Results and Discussion 119
7.7 Summary 121
7.8 Appendix (Chapter 7) 123
Chapter 8 131
Conclusion and Future work 131
8.1 Conclusion 132
8.2 Future Work 136
8.3 Work in Progress 137
References 139

List of Figures

Figure 1- 1: Shows an example of US scan images of ovarian tumour. 2
Figure 1- 2: General framework for the proposed ovarian tumour diagnostic. 6

Figure 2- 1: Few imaging modalities (a) MRI knee (b) chest CT scan (c) US ovarian scans (Medical Image Modalities 2019). 11
Figure 2- 2: (a) Example of Ultrasound Machine (b) Different Types of US Probes. (Gurjar 2017).....12
Figure 2- 3: Sender /Receiver in Ultrasound probe (D.Wong 2011).....13
Figure 2- 4: Illustration of the transverse, sagittal, and coronal planes of the body. (Bodytomy 2019). 13
Figure 2- 5: The Female reproductive system (anatomy 2019). 14
Figure 2- 6: Normal and Cancer cells (Media 2019). 15
Figure 2- 7: Histological types of ovarian tumours. 16
Figure 2- 8: Illustration of the extent of the spread of ovarian cancer - (a) - (d) for stages 1 to 4 (UK Cancer Research 2018). 18
Figure 2- 9: (a) Ultrasound Image, (b) Fan Area and a Margin Area. 19
Figure 2- 10: Block diagram of the major steps of CAD. 20
Figure 2- 11: The usual tissue model in ultrasound imaging. 23
Figure 2- 12: Diagram Representation of the Principle of SVM. (a) SVM attempts to Maximise the Margin from the Hyperplane to find the Best Separate between two Classes (red positives from blue negatives) (b) Optimal separating Hyperplane. 27
Figure 2- 13: Illustration of the working of kNN classifier. 29

Figure 3- 1: Illustration of speckle noise and its effect: (A) Example of speckle noise and (B) US image of an ovarian tumour corrupted by speckle noise. (Dangeti 2003)...33
Figure 3- 2: Manually segmented ovarian tumour regions marked by yellow borders. 40
Figure 3- 3: Histograms of ultrasound ovarian tumour image (a) Benign, and (b) Malignant. 41
Figure 3- 4: LBP histograms for ultrasound ovarian tumour image of (a) Benign, (b) Malignant. 43

Figure 4- 1: (a) Single LBP code for 3x3 block (P=8, R=1), (b) Image with its LBP transform. 52

Figure 4- 2: Different texture primitives detected by LBP. 53
Figure 4- 3: HOG process. 55
Figure 4- 4: FD example (scale factor (1/r) and number of boxes (N(r))). 56
Figure 4- 5: Hu's moments. 58
Figure 4- 6 : Performance measures obtained using the majority voting fusion using SVM & kNN (k value =1) classifiers. 65

Figure 5- 1: (a) A US image of an ovarian tumour corrupted by different levels of speckle-noise: Low level in part (b), High level in part (c). 69
Figure 5- 2: Kourtosis & Skewness of tumour tissue images (A) little solidity; (B) predominantly solid. 72
Figure 5- 3: Model1: Speckle noise based on the whole image. 74
Figure 5- 4: Ovarian tumour samples illustrate the different level of speckle-noise 75

Figure 6- 1: (a) Original ultrasound ovarian tumour image (b) its DFT spectrum (c) spectrum phase 89
Figure 6- 2: Samples of FFT spectrum images for Benign and Malignant tumour RoIs 90
Figure 6- 3: The process of FFGF features 91
Figure 6- 4: An example of Morphologically operation: (a) Bini image (b) image after closing the objects 93
Figure 6- 5: Classification results based on FFGF using SVM & kNN classifiers 94
Figure 6- 6: The block diagram of FFT Spectrum texture-based classification 98
Figure 6- 7: Display example of LBP image using the FFT spectrum 98
Figure 6- 8: shows the results using SVM classifier 100
Figure 6- 9: Demonstration the results using kNN classifier 101
Figure 6- 10: Features extraction methods in the LBP transformed domain 102
Figure 6- 11: Impact of normalising histograms in the LBP domain 104

Figure 7- 1: Interface (a) import image, (b) mark RoI by expert and crop by software and (c) output a class prediction. 109
Figure 7- 2: Performance of Version 1, based on 7 schemes without pre-processing for prospective tests 111
Figure 7- 3: Performance of Version 2, 9 schemes without pre-processing 112
Figure 7- 4: Performance of Version 2, 9 schemes with adaptive de-noising 113

Figure 7- 5: Explanation of the general structure of a CNN.....116
Figure 7- 6: Illustrations an example of data augmentation.....118
Figure 7- 7: Training progress using Resnet50 without pre-processing.....120
Figure 7- 8: Training progress using Resnet50 with pre-processing.....120
Figure 7- 9: Testing results using Resnet50 with and without pre-processing.....121

Figure 8- 1: Different US images (a) 3D scan of benign tumour (b) 2D colour (c) Doppler Image.....137
Figure 8- 2: Shows the results for the brest cancer using FFGF (Minor axis length).138

List of Tables

Table 3- 1: Lists some examples that illustrate the different problems mentioned above. 39

Table 4- 1: Histopathology of ovarian masses 59
Table 4- 2: Classification results based on SVM classifier.....61
Table 4- 3: Classification results based on kNN classifier 63

Table 5- 1: Performances of 7 texture feature methods with blind pre-processing and without pre-processing based on SVM & kNN (k value =1) classifiers.....70
Table 5- 2: Samples of a class 1 and class 2 RoIs of ovarian ultrasound images. 74
Table 5- 3: Performance of 5 texture-based diagnostic schemes post model 2 adaptive de-noising based on SVM and kNN (k value =1) classifiers. 78
Table 5- 4: Performance of different Majority Voting fusion schemes based on SVM and kNN (k value =1) classifiers.....79

Table 6- 1: FFT and binary spectrums of ultrasound images of benign and malignant ovarian masses. 95
Table 6- 2: Classification result based on Minor, Major and Area using SVM & kNN based on (k value =1) classifiers. 96
Table 6- 3: Performance of Frequency domain texture-based diagnosing schemes based on SVM and kNN (k value =1) classifiers. 99
Table 6- 4: Performance of a combination of texture features extracted from LBP domain based on SVM and kNN (k value =1) classifiers. 102

Table 7- 1: Presents the format of saving decision. 110
Table 7- 2: Display version 1 of the prospective test based on 7 Extraction Methods without Pre-processing. 123

Table 8- 1: Samples of FFT spectrum images for breast cancer. 138

Abbreviations

CA-125	Cancer Antigen 125
CAD	Computer Aided Diagnosis
B-mode US	Brightness-mode Ultrasound Imaging
ROI	Region of Interest
US	Ultrasound
SVM	Support Vector Machine
kNN	k-Nearest Neighbour
IOTA	International Ovarian Tumour Analysis
LBP	Local Binary pattern
ULBP	Uniform Local Binary Pattern
HOG	Histogram of Orientation Gradient
FD	Fractal Dimension
GS	Gestational Sac
YS	Yolk Sac
FFGF	Fast Fourier-based Geometric Features
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
L1O	Leave -one-out
Stdev	Standard Deviation
FN	False Negative
FP	False Positive
TN	True Negative
TP	True Positive
CNN	Convolutional Neural Network
DL	Deep Learning

Chapter 1

Introduction

This thesis follows the recent trend in benefiting from advances in machine learning and artificial intelligence for digital health and in particular for the automatic analysis of medical images for computer-aided diagnostic systems. In this chapter, we first give an overview of the focus of our research investigations regarding the texture of ovarian ultrasound images. Then, we describe the motivating of this research, as well as the aim and objectives of this research. After that, the general framework of the thesis for proposing ovarian tumour classification using 2-D statics images will explain. Finally, the list of contributions and publications of this research.

1.1 Overview

The tissues and organs of the human body are made up of tiny building blocks, which are named cells. These cells are divided into two parts: normal and abnormal cells. The cells in the healthy body divide regularly, grow to make new cells, and die in an orderly way. However, cells start to divide and split into body tissues/organs in an uncontrolled manner; the cells are called abnormal or cancerous cells. There are different kinds of cancer, such as breast, colon, ovarian cancer, bowel, etc. (American Cancer Society 2018).

Ovarian cancer is the deadliest malignancy of the female reproductive system (American Cancer Society 2018). Moreover, it is a disease when part of cells in the ovary develops and changes in an out-of-control manner to become cancerous cells. The ovarian cancer is called the “silent killer” because it can be spread and developed before women are even aware that they have it, i.e. there are not specific symptoms of the ovarian carcinoma until getting to the advanced stage of cancer (Gajjar, et al. 2012). However, successful treatment of ovarian cancer, and any type of cancer, greatly depends on early detection and diagnosis. Early detection tools need to be established to meet the unmet needs of ovarian cancer patients and save their survives. There are different methods to identify ovarian cancer, such as medical ultrasound imaging, laparoscopy and a blood test known as (CA 125) to test the level of tumour marker and others (American Cancer Society, 2018; Togashi, 2003). However, it is essential to distinguish between different categories

of ovarian cancer, for example, benign from malignant, not only to ensure the suitable treatment for malignant cases but additionally to avoid unnecessary diagnostic processes such as surgery for the non-malignant cases (Valentin, et al. 2006).

This research is concerned with investigating and developing digital tools and techniques for the analysis of ultrasound scan images to facilitate the detection and classification of ovarian cancer (see Figure 1.1). Ultrasound scanning is the highest modality used in hospitals to distinguish between different ovarian tumours and also it is the prime triage method before treatment (Kinkel, et al. 2000). Even, however, it cannot essentially prevent the surgery which can help to narrow down the different analysis and determining the level of suspicion for malignancy, mostly together with the serum CA-125 level (Togashi 2003)**.** Nevertheless, the performance of ultrasound imaging based on the morphology assessment is limited because of the significant number of false-positive results (Kupesic and Plavsic 2006) and increases the issue of how to correctly interpret the images (Gramellini, et al. 2008). According to (Fishman, et al. 2005) the grey-scale scoring systems to detect ovarian cancer can result in many unnecessary operations. This is due to the limitations in the human eye visual system reader tiredness, interruption, and the most important is the overlapping between different images structures, which can camouflage disease in images, may cause errors in clarification. (Giger, Chan and Boone 2008). Properties of an image can be quantified and measured through a technique called texture analysis, which analyses the pixel position and intensity. According to (Mayerhoefer, et al. 2008), the process of texture analysis can be applied to ultrasound images through the use of a computer. The computer is able to recognize specific anatomical structures in the image through mathematical patterns identified in the grey level distribution of the pixels in the image.

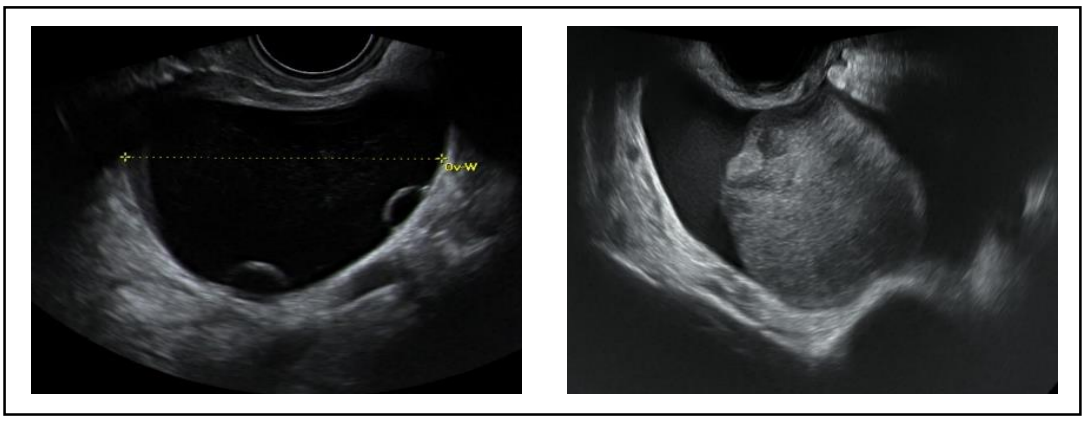


Figure 1- 1: Shows an example of US scan images of ovarian tumour.

The technique of texture analysis is used extensively in the medical field, particularly in the past two decades. In addition, this technique is also used in other applications, such as remote sensing, etc. In the medical field, the output obtained from the image texture analysis can be inputted into a program that can assist in diagnosis, called the computer-aided diagnosis (CAD). CAD is often used to improve the sensitivity and specificity of radiographic images (Dhawan, Atam P 2011). An important advantage of using CAD is that a computer method can reduce subjective interpretation, allowing for repeatability of results (Smyth, et al. 1997). The texture analysis technique can evaluate pixel intensity variation that refers to a specific physical attribute or variation in the image.

Accordingly, the ultrasound images with texture variation (Davis 1980) , particularly in the medical context, reveal the overall structure of cells or tissues (Szczypinski, et al. 2009), either with or without pathological changes (Mayerhoefer, Breitenseher, Amann, & Dominkus, 2008; Xian, 2010). Using the texture analysis, the interpretation of the ultrasound imaging results is primarily based on the underlying basis that abnormal cells or tissues produce distinct texture variation or intensity (modified ultrasound signal) from those of normal cells or tissues (Morris 1988). The texture of an ultrasound image can be defined as the smoothness and structure of the objects. The texture is critical as it is a primary characteristic of an image that is used for the analysis and interpretation of ultrasound images (Kurani, et al. 2004). Additionally, it is conducted to compute the texture and properties of the components in the image (Mathias, Tofts, & Losseff, 1999; Nailon, 2010), as well as to define critical textural characteristics for an ROI. Texture analysis is typically used in the medical field to help distinguish between normal, healthy regions, and regions that require treatment. It can also be used to differentiate between various tissues and organs (Castellano, et al. 2004).

1.2 Research Motivation

Ultrasound ovarian images have been playing an increasingly important role in the detection and diagnosis of the types of tumour, which is a crucial factor for the effective treatment of patients. The current manual examination system imposes a big challenge for most specialised medical centres due to limited numbers of expert radiologists needed to analyse a large number of various cases. In such situations, the possibility of misdiagnosis of abnormality/normality of ovarian tumours with tragic consequences could not be ruled out.

On the other hand, early detection of ovarian cancer, or any type of cancer, remains a challenging task. Ultrasound has demonstrated utility in detecting an ovarian tumour in

asymptomatic women, also can help differentiate benign from malignant lesions. However, inexperienced ultrasound operators are always meeting difficulties in distinguishing between different kinds of tumours, which eventually lead to a lower rate of true diagnosis. Incorrect diagnosis may lead to unnecessary biopsies/surgery, or worse, missed cases. Automating certain aspects of ultrasound image analysis can help to provide supporting tools to help increase their diagnostic accuracy. This is the main motivation for the work reported in the thesis as part of a wider related research work conducted by the Buckingham biomedical image analysis team.

A computer-based system for categorisation of ultrasound images could contribute to improved performance of decision support systems for the diagnosis of ovarian tumours. Such a computer-based system benefits from an interdisciplinary technology that combines image processing methods and experts' knowledge for greatly, better accuracy of abnormality detection, hence greatly decreasing the false-negative rate and improving the true positive rate.

1.3 Research Aim and Objectives

In a medical diagnosis of ovarian cancer, medical experts examine ovarian ultrasound images to visually identify any characterising features of suspected tumours using some measurements in addition to clues from the tissue texture according to well established medical practices. A group of medical experts with extensive technical knowledge and experience collectively assess and finalise their findings of the ultrasound imaging results. In this study, we attempt to encapsulate, as much as possible, the knowledge of domain experts for designing and developing CAD tools through the implementation of machine learning strategy. In particular, the intended CAD tools are meant to be used for the gynaecological ultrasound imaging in detecting ovarian cancer and are therefore developed through the rigorous process of training and testing using ovarian ultrasound images with hybrid image features extracted automatically from different image domains of representation, e.g. spatial and frequency domains.

Based on the knowledge gained from domain experts, we estimate and categorise the image texture information with the objective of establishing a mathematical model suitable for analysis by machine learning schemes. Most importantly, it was expected that the included features were either challenging for the human visual system or visually inaccessible for medical experts (such as features from the Fourier spectrum). Accordingly, we may need to incorporate a diverse set of image features as well as fuse

them at different levels to evaluate the diagnostic performance of developed the automated computerised machine learning tools.

Following the feature extraction process using computational techniques, this study also performed feature quantification and analysis to identify any notable anatomical and textural changes in a tumour. The extracted features were then used as distinguishing attributes between different classes. The developed tools in this study were expected to significantly assist the medical experts in the decision-making process for feature classification cases at a specific level of confidence. These following points summarised the main objectives of our work:

- ✓ To identify and define texture features to be extracted from ultrasound images in spatial, frequency and transformed domains followed by finding suitable algorithms that are capable of recognising different types of tumours with a high level of accuracy. The different mathematically defined texture features are expected to extend the list of the features beyond those that manually extracted by medical experts.
- ✓ To design and develop ultrasound ovarian tumour image classification schemes based on the texture features, identified above, and evaluate their performances as well as those of fused combinations of them using appropriate datasets based on standard classifiers trained and tested according to standard experimental protocols.
- ✓ To determine factors that influence the performance of different feature scheme in uncontrolled conditions. For example, we determine the impact of speckle noise in ovarian ultrasound images on the performance of the selected texture features in distinguishing different types of ovarian tumours. We then investigate various speckle de-noising schemes and design effective and adaptive solutions to suppress this type of noise if and when needed.
- ✓ To exploit the gained knowledge from the realization of the above objectives and build the intended CAD software tools that can be integrated within a clinical framework. Since these tools are meant to support the clinician through the decision- making the process for diagnosis of ovarian tumours, we should attempt to test the performance our tools in a clinical setup through a prospective test.

1.4 Research Framework of the Proposed Ovarian Tumour Classification

Computer-based analysis plays a major role in ultrasound image analysis and could provide high levels of accuracy as well as help the sonographer to obtain accurate results for better diagnosis. This section will explain the Computer-aided diagnostic of the proposed automatic image diagnostic, which can be shown in Figure 1.2: -

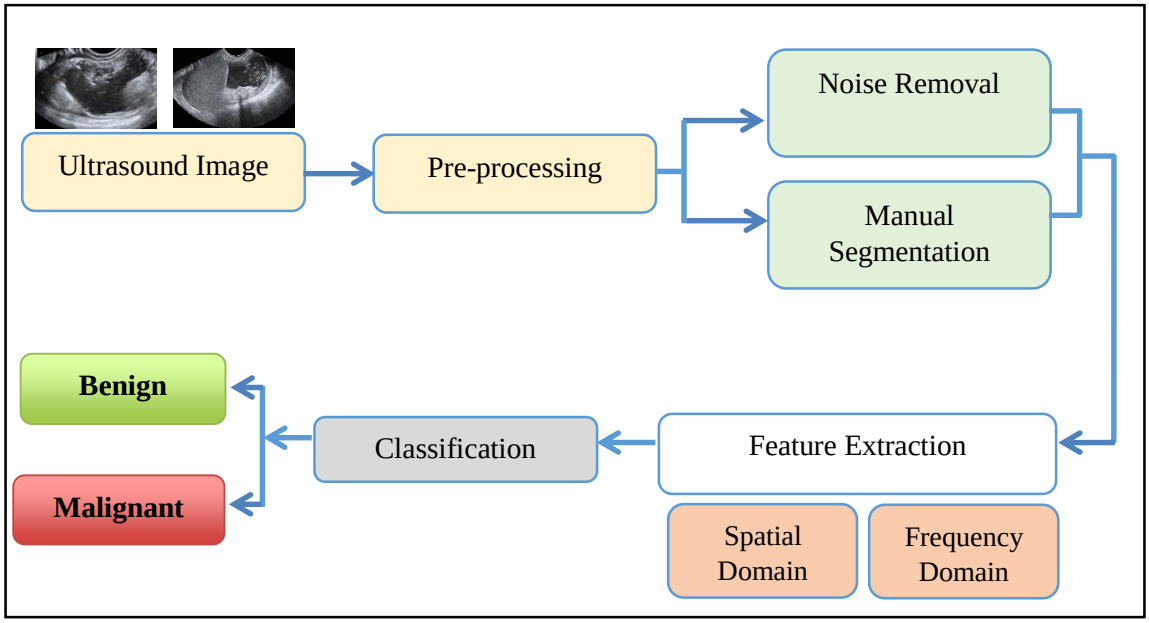


Figure 1- 2: General framework for the proposed ovarian tumour diagnostic.

- ✓ Stage One (Pre-processing): - For the duration of data acquisition processes, the ultrasound images will affect by random noise. This type of noise is an inherent property of medical ultrasound imaging, and because of this noise, the image resolution and contrast become reduced, which affects the diagnostic value of this imaging modality. Therefore, noise reduction is an essential pre-processing step to highlighting the details of different objects inside the region of interest.
- ✓ Stage Two (Segmentation): Once we suppressed the level of the speckle noise, the ROI should be extracted from the image background. In fact, there are many techniques to segment the ROI, such as thresholding, region growing, etc. However, it has been mentioned in the literature that the segmentation of ovarian ultrasound images is still a complex, challenging, and it is an unsolved problem due to different issues.
- ✓ Stage Three (Feature Extraction): - This stage is aimed to extract the most important texture features from the segmented images of ovarian tumours. Various feature extractions methods are used in both spatial and frequency

domains. These texture features are additional and different from radiologists features which have been extracted manually.

- ✓ Stage Four (Classification): - this is the last stage in our work, which is fed the extracted features from ovarian ultrasound images to the classifier to make a diagnostic verdict. In this research, we are focused on binary classification, which is Benign and Malignant ovarian tumours. Every stage of our work will be described in detail in the following chapters.

1.5 Thesis Contributions

The following is a summary of the main contributions achieved in this research to overcome the limitations of existing ultrasound pre-processing and texture analysis techniques.

1. Evaluating the Power of the state-of-art texture-based features alone to Discriminate Benign from Malignant Ovarian Tumours using ultrasound images. We illustrate and analyse the effectiveness of 7 texture techniques without pre-processing. Then, the simple majority fusion method was applied to an odd number of texture features to improve classification accuracy.
2. Developing an adaptive system to suppress the level of speckle-noise and enhancing the texture of Benign and Malignant cases. This adaptive system approach is applied pre-processing only when it is needed (i.e. local pre-processing instead of globe pre-processing). The new adaptive system has improved the performances of the texture features.
3. Proposing a new set of features based on the spectrum of Fast Fourier based Geometric Features FFGF. In this work, we modified on the exciting FFGF by applied our new adaptive pre-processing as well as using different binarisation method to extract the bright elliptical shape from the centre of the FFT spectrum.
4. Proposing a new combination of feature extraction techniques based on the image spectrum of the FFT a new approach. In this approach, the spectrum of the FFT is used instead of binarized the central of the spectrum of FFT since it has a rich source of texture information. After applying the 7 texture feature methods based on the FFT image, we found that it outperforms their spatial domain counterpart.
5. Proposing a new set of texture features using LBP transform domain, i.e. instead of using the histogram as features from the LBP, we used the LBP image as input image then applied another feature extraction technique on the LBP image.

6. Developing software for Prospective clinical evaluation of texture-based features analysis of US ovarian scans for discriminating Benign and Malignant tumours.

1.6 List of Publications

i. Dheyaa Ibrahim, Hisham Al-Assam, Hongbo Du, Jessica Farren, **Dhurgham Al-karawi**, Tom Bourne, Sabah Jassim; **Automatic segmentation and measurements of gestational sac using static B-mode ultrasound images**. Mobile Multimedia/Image Processing, Security, and Applications. International Society for Optics and Photonics, USA, 2016, vol.9869.

ii. **Dhurgham Al-karawi,** Ahmed Sayasneh, Hisham Al-Assam, Sabah Jassim, Page N, D Timmerman, Tom Bourne, Hongbo Du; **An automated technique for potential differentiation of ovarian mature teratomas from other benign tumours using neural networks classification of 2D ultrasound static images: a pilot study.** Mobile Multimedia/Image Processing, Security, and Applications. International Society for Optics and Photonics, USA, 2017, vol.10221.

iii. **Dhurgham Al-karawi**, Chiara landolfo, Hongbo Du, Hisham Al-Assam, Ahmad Sayasneh, Dirk Timmerman, Tom Bourne , Sabah Jassim; **Prospective clinical evaluation of Texture-based Features Analysis of Ultrasound Ovarian scans for Distinguishing Benign and Malignant Tumors.** Australasian Journal of Ultrasound in Medicine. Wiley Online Library.2019, vol.22.

iv. **Dhurgham Al-karawi**, Hisham Al-Assam, Hongbo Du, Ahmad Sayasneh, Chiara Landolfo, Dirk Timmerman, Tom Bourne, Sabah Jassim; **An Evaluation of the Effectiveness of Image-based Texture Features Extracted from Static B-mode Ultrasound Images in Distinguishing between Benign and Malignant Ovarian Masses.** It's been submitted to IET Image Processing Journal.

v. **Dhurgham Al-karawi**, Chiara landolfo, Hongbo Du, Hisham Al-Assam, Ahmad Sayasneh, Dirk Timmerman, Tom Bourne , Sabah Jassim; **A machine-learning algorithm to distinguish benign and malignant adnexal tumours from ultrasound images.** 29th **World** Congress on Ultrasound in Obstetrics and Gynecology, Berlin, Germany, 2019.

1.7 Dissertation Layout

The rest of the thesis is structured as follows: -

Chapter 2: Presents a background knowledge about the medical images, and more details relevant to ultrasound image analysis, as well as highlights the current clinical techniques used to characterise ovarian tumours. Furthermore, it will explain a computational background of different methods of image processing which are related to this thesis.

Chapter 3: This chapter presents the literature review of existing works on speckle-noise reduction techniques for ultrasound images. Also, reviews the methods on extracting regions of interest in medical ultrasound images. Moreover, it will review the current works on texture-based features for ovarian tumours classification using ultrasound images.

Chapter 4: This chapter provides the development of several image texture-relevant features based on spatial and frequency domains to categorise between benign from malignant ovarian tumours. As well as, the datasets used in this research and experimental protocols.

Chapter 5: This chapter is divided into two parts; the first part is indicated the proposed new approach of speckle-noise reduction. The second part is proposed the development of a procedure to mitigate the second factor on variation in ROI sizes.

Chapter 6: This chapter is described as a novel idea based on the spectrum of fast Fourier transform to classify binging from malignant ovarian tumours. As well as, we will explain our new texture feature based on different domains.

Chapter 7: Presents the software for the prospective clinical test based on hand-crafted features in comparison with Deep learning tools.

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

Chapter 2

Medical and Computational Background

The multidisciplinary nature of this research project adds to the challenge of writing up the thesis that reports the contributions made and the knowledge uncovered in a manner that is accessible to researchers and practitioners from the different disciplines. To meet the accessibility requirements, this chapter is entirely devoted to briefly explaining the relevant background materials from the participating disciplines insufficient details to enable readability of the rest of the chapters. In section 2.1 and 2.2, we begin with a brief description of general medical imaging systems before introducing various types of ultrasound imaging systems and their use for diagnosing gynaecological abnormalities. Section 2.3 is given a general discussion about the technology fields of image processing/analysis, and machine learning will form the second part of the chapter. Finally, the chapter will focus on describing the technical aspects of developing automated image-based medical diagnostic systems in general, highlighting the ingredients and requirements of our intended system of ovarian tumour diagnostic from Ultrasound Ovarian scans.

2.1 Medical Imaging Systems

Medical imaging is a discipline within the medical field that mostly involves the use of various radiology technologies to help the clinics visualise human body organs/tissues for detection of medical abnormalities. Medical imaging is sometimes referred to as diagnostic imaging, due to the fact it is frequently used by the doctors in conjunction with other types of tests and procedures to arrive at a highly reliable diagnostic decision. However, the use of radiological scan images as a complementary aiding tool is by no mean limited to diagnosing different diseases but extends to a variety of healthcare objectives that include but not limited to assessing the degree of success of treatments/surgery, understanding/interpreting functionalities of organs/tissues, and routine population screening. There are different types of medical image modalities that are used to access and monitor mostly internal human body organs/tissues hidden behind skins, bone structures, and fat tissues. Although digital and thermal cameras are also used for medical examinations, radiology imaging is based on different principles that require the use of specialised devices/procedures to create/display images susceptible to digital

image processing/analysis. By definition, radiology imaging is based on using certain waveform sub-bands of the electromagnetic spectrum depending on the nature of the organ/tissue to be scanned. The list of radiography imaging schemes is growing, but the most common modalities are X-Ray Computed Tomography (CT) also referred to as Computed Axial Tomography (CAT) scans, Magnetic Resonance Imaging (MRI), Ultrasonography, and Nuclear medicine functional imaging modalities such as MRI and Positron Emission Tomography (PET). There are various methods to obtain and process these types of images, depending on the deployed equipment and on the nature of the scanned body region. In CAT scan, beams of X-rays spin around the target (e.g. internal organs, bones, blood vessels) from several angles and special sensors then detect the output signal after penetrating the object to form an image of the scanned object (Dougherty 2009). Instead of using radiation, the MRI modality uses a combination of magnetic fields and radio waves to generate images of the scanned body anatomy and their physiological processes. The US modality relies on sound waveforms to create a "cross-sectional" view of anatomical structures. Figure 2.1 shows the output from several kinds of medical image modalities.

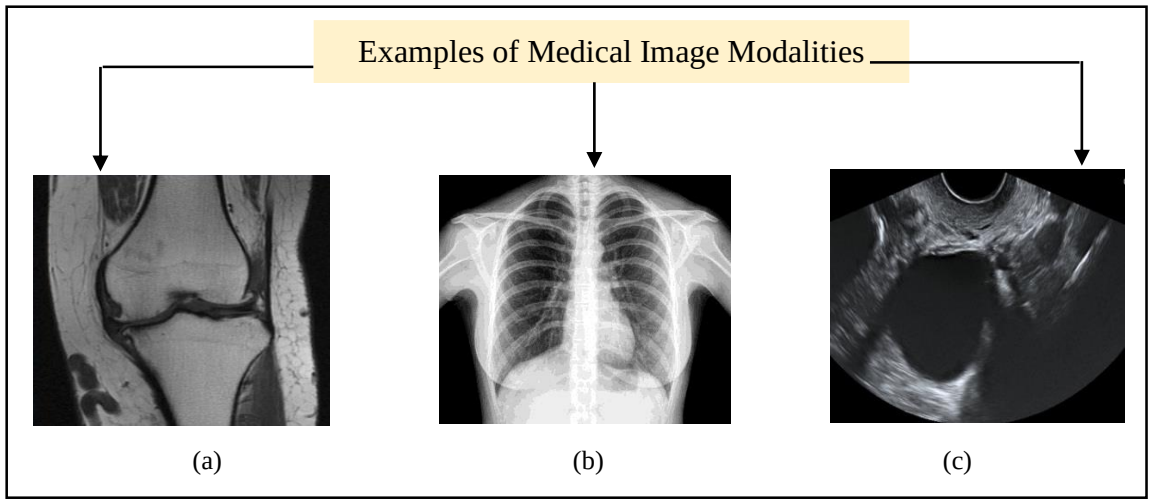


Figure 2- 1: Few imaging modalities (a) MRI knee (b) chest CT scan (c) US ovarian scans (Medical Image Modalities 2019).

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 the US as a medical analytical tool to image the brain (Edler and Lindstrom 2004). Nowadays, US is one of the most commonly used imaging technologies in medicine, especially in gynaecological abnormality/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, for example, MRI and CT (Hoskins, Martin and Thrush 2010). 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).

2.1.2 The Equipment of Ultrasound

An ultrasound scanner involves of several 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 comes in different shapes and sizes for use in different scanning purpose. The probe sends out a range of frequency sound waves into the body and afterwards listens for the returning echoes from the tissues in the body. The ultrasound image is directly visible on a video show screen (monitor). The image is created based on the amplitude (strength), frequency and time it yields for the sound signal to return from the area of the patient being examined to the transducer and the kind of body structure the sound travels through (Hoskins, Martin and Thrush 2010). 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 2.2 shows three different categories of the ultrasound machine.

(a) **(b)**

Figure 2- 2: (a) Example of Ultrasound Machine (b) Different Types of US Probes. (Gurjar 2017)

2.1.3 Scanning Process

Prior to scanning, a thin layer of jelly is applied on the probe and the skin. This layer helps to ensure that all the sound waves go into the body. The probe is composed of a transmitter and a receiver. While the transmitter emits an ultrasound pulse, the receiver collects the reflected pulse from tissues and organs within the body. The ultrasound machine then analyses the time it takes for the pulse to return (Hoskins, Martin, & Thrush,

2010; Powers & Kremkau, 2011). The scanning procedure is depicted in Figure 2-3 below.

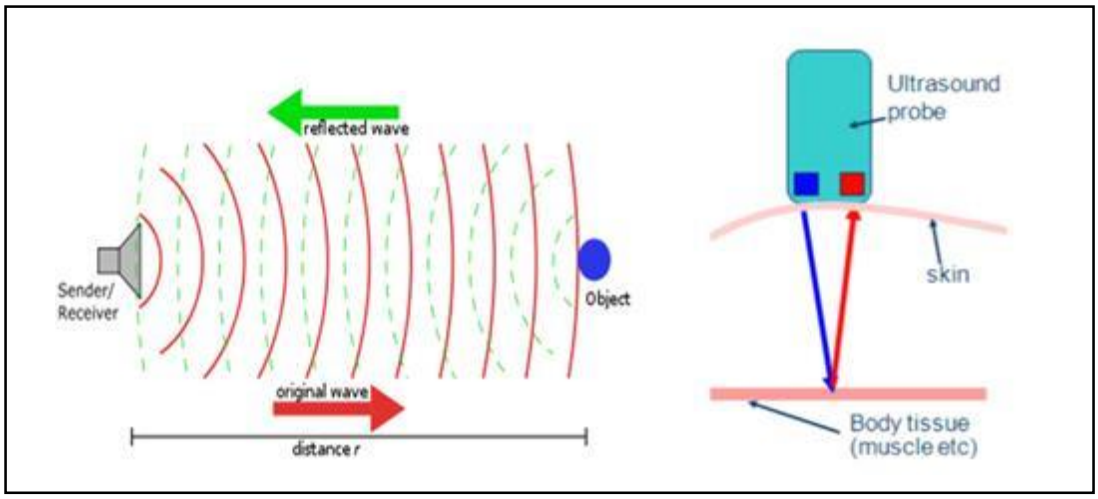


Figure 2- 3: Sender /Receiver in Ultrasound probe (D.Wong 2011)

The most common type of ultrasound image is named Brightness-Mode (B-mode). It is a two-dimensional sequence which is crossed section if images of the scanned part (Hoskins, Martin and Thrush 2010). The real-time scanning could be taken in various planes as described below and shown in Figure 2-4. However, there are more types of ultrasound images, which are 3D, 4D, as well as Doppler. However, in this research, we are focusing on 2D ultrasound images.

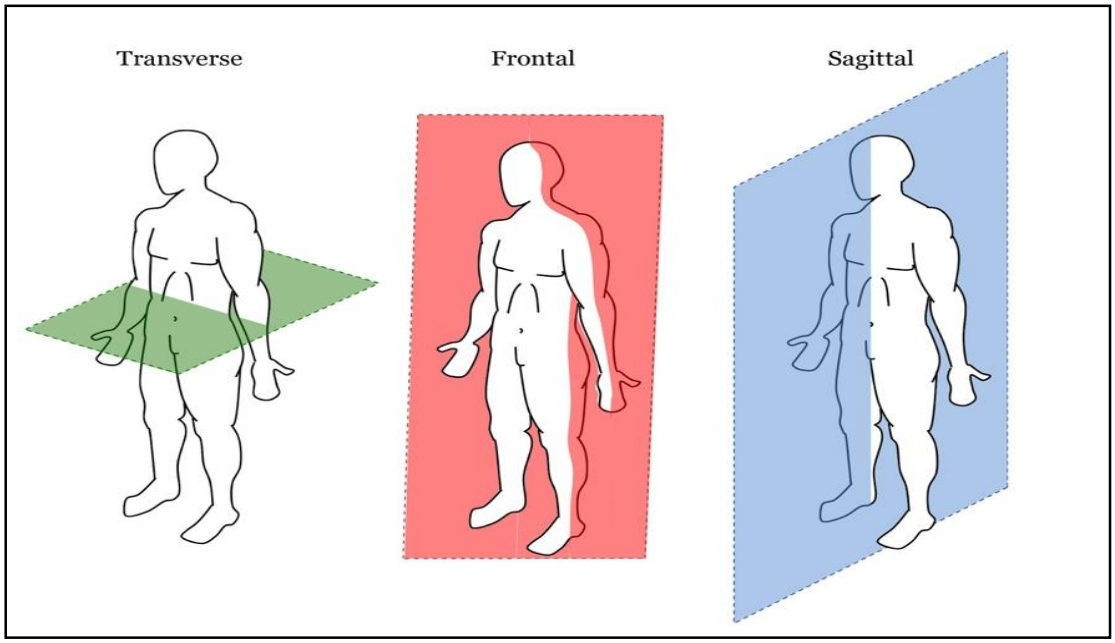


Figure 2- 4: Illustration of the transverse, sagittal, and coronal planes of the body. (Bodytomy 2019).

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

More details about the process of generating US scan images and the various devices deployed in this process can be found in (Hoskins, Martin and Thrush 2010). But now we will turn our attention in the next section to the use case of interest in this thesis, i.e. in the medical field of gynaecological.

2.2 Background of Human Female Reproductive System

The female reproductive organ consists of five parts: (vagina, uterus, fallopian tubes, Cervix, and ovaries), and each part has a different role, (Hamlett and Koob 1999). Figure 2-5 is a diagram illustration of the different parts of this reproductive system.

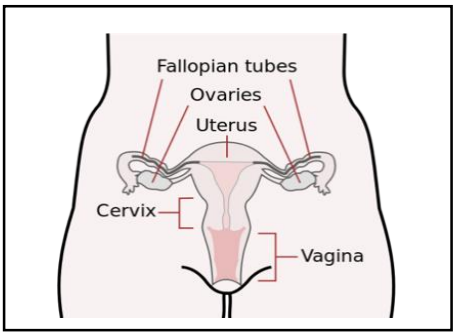


Figure 2- 5: The Female reproductive system (anatomy 2019).

The next section is to brief of the description of these 5 components of the reproductive system.

i. ***Fallopian tubes***: a couple of muscular tubes that extend from the left and right corners of the uterus to the edge of the ovaries.
ii. ***Ovaries***: There are two ovaries, one on each side of the body. These ovaries contain ova inside follicles for maturation through the woman's reproductive life.
iii. ***Uterus***: This part is responsible for feeding the fetus until birth.
iv. ***Cervix***: This part is connecting tube between the uterus and the vagina.
v. ***Vagina*** is a flexible, muscular tube connecting the cervix to the outside body part.

In this thesis, we are focusing on automatic computerised ovarian tumour diagnosis, i.e. digital schemes for diagnosing abnormalities in the ovary parts of the female reproductive system by analysing 2D B-mode ultrasound images of the ovaries. The female reproductive has two ovaries, one on the left and right side of the uterus in the pelvic

region. Ovaries are responsible for producing the ova and the female sex hormones, such as (oestrogen and progesterone). These hormones control the development of the female body, including the body shape, the menstrual cycle, pregnancy, breasts and body hair grow. The shapes of the ovaries are ovoid, of length in the range from 3 to 5 cm and weight between 2 to 4 grammes (Hamlett and Koob 1999).

2.2.1 Overview of Ovarian Cancer

Cancer starts when the cells in a part of the human body become out of control; i.e. become abnormal cells. There is a difference between the growth of cancer cells and normal cells. Instead of dying after a certain number of division rounds, cancer cells continue to divide, forming new networks of irregular cells that in some cases are referred to as a tumour. Besides, cancer cells can attack (grow into) other tissues, something that uncommon in the healthy cells.

Emerging Systems Biology, evidence demonstrates that cancer cells actively rewire cellular networks during carcinogenesis (Modos, et al. 2017). Subsequently, uncontrolled growth of the irregular cells and invading other tissues may result in cancerous mass (Gajjar, et al. 2012). Figure 2-6 illustrations the normal and cancer cells

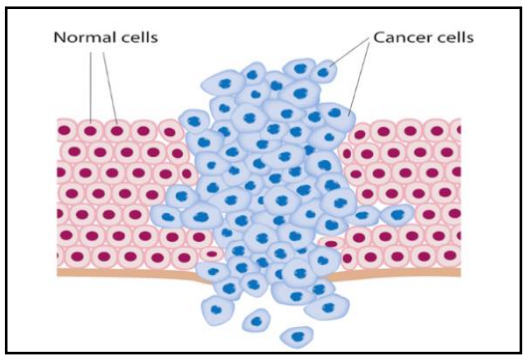


Figure 2- 6: Normal and Cancer cells (Media 2019).

In general, the division of any cell in the body is controlled by a certain gene. If this gene misses its function, it might lead to the 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 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 also spread via the bloodstream or the lymph system and reach to other parts of the body (Tan, Agarwal and Kaye 2006).

Ovarian cancer is strongly related to age. Although it could grow at any age, it is most likely to occur in older women, especially from age 55 and over (American Cancer Society 2018). According to (Bell 1991) epithelial tumours occur in pre-menopausal and

post-menopausal age, and 85% of ovarian cancer cases are epithelial cell (Chan and Perlas 2011). Through extensive study has carried out, the real reasons for ovarian cancer are still unidentified. The two most effective factors linked to the risk of rising ovarian cancer are age and the presence of certain gene mutations. Further factors that can affect the risks of increasing ovarian cancer are described below:

- ✓ Infertility, the risk is lower in women who had birth, likened to women who did not have children (Permuth-Wey and Sellers 2009) .
- ✓ The history of the family, women with a mother or sister detected with ovarian cancer, have a higher risk of developing ovarian cancer. Likewise, women with previous breast cancer have double the risk of ovarian cancer (Permuth-Wey and Sellers 2009).

2.2.2 Histological Types of Ovarian Tumours

Ovarian carcinoma is not a single disease, and there are more than 30 types and subtypes of ovarian tumours owned histological (diseased tissue) form and biologic behaviour. Based on that, the panel experts classify the ovarian tumours into three groups according to the kind of cells/tissue (Scully and Sobin 1999), Figure 2-7 shows histological types of ovarian tumours

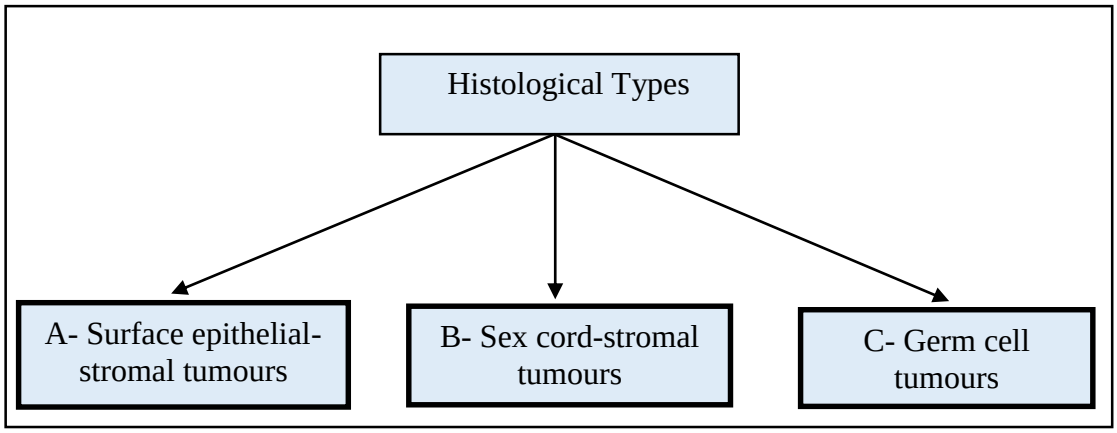


Figure 2- 7: Histological types of ovarian tumours.

A-Surface epithelial-stromal tumours: - Tumours compose of more than one types of epithelium and stromal cells in a variety of combinations. Furthermore, these tumours arise from the surface epithelium of the ovary, or it is derivatives (the surface epithelial implication glands and the adjacent ovarian stromal). The surface epithelial-stromal tumours are classified according to the tumour type (mucinous, endometrioid, serous, clear cell, transitional cell) and further sub-classified as benign, borderline or malignant (Scully and Sobin 1999).

B. Sex cord - stromal tumours: - this is the second group of the ovarian tumours. The sex cord - stromal tumours are accounted for by approximately 8% of all ovarian tumours and around 7% of all malignant ovarian tumours (Chen, et al. 2003). Fibromas and fibroatheromas tumours are the sub-types of the sex cord-stromal tumours, which are uncommon benign tumours of stromal origin. These types of tumours are accounted for by only 4% of all ovarian tumours but characterise the most common solid primary cancers of the ovary. However, Fibromas and fibroadenomas arise from spindle mesenchymal cells that produce collagen and can be associated with Meigs syndrome. Compared to Fibromas, fibroatheromas have a small population of theca cells that include intracellular lipid and potential to show estrogenic activity. (Yen, et al. 2013).

C. Germ cell tumours: - the germ cell tumours are the last group of the ovarian tumours histological. They are accounted for by 20-25% of all ovarian neoplasm however around 3 % of these are malignant. The most common type tumour in this group is Teratoma as a benign tumour that rarely undergoes malignant degeneration. However, the most malignant ovarian germ cell tumours are composed of primitive or immature elements. (Tangir and Schwartz 2003).

2.2.3 Stages of Ovarian Cancer

There are four stages in ovarian cancer, generally determined by the location of the cancer cells (Kurtz, et al. 1999). In the first stage, also known as early cancer, only the ovaries may contain cancerous cells. In stage 2, the cancerous cells are either in the ovaries or the fallopian tubes, with the possibility of cells spreading to other tissues in the pelvis, like the uterus. In stage 3, the cancer is more widespread, with cancer cells found outside the pelvic area and into the abdominal area. At this stage, it is typical for the size of the tumours to be larger than in stage 1 or 2. When cancer cells spread further into other organs, such as the lungs or the liver, this is considered to be stage 4. Even if the other organs contain cancer cells, other than those originated from the ovaries, stage 4 masses are still identified as ovarian cancer. Figure 2-8 illustrates the extent of the spread of ovarian cancer.

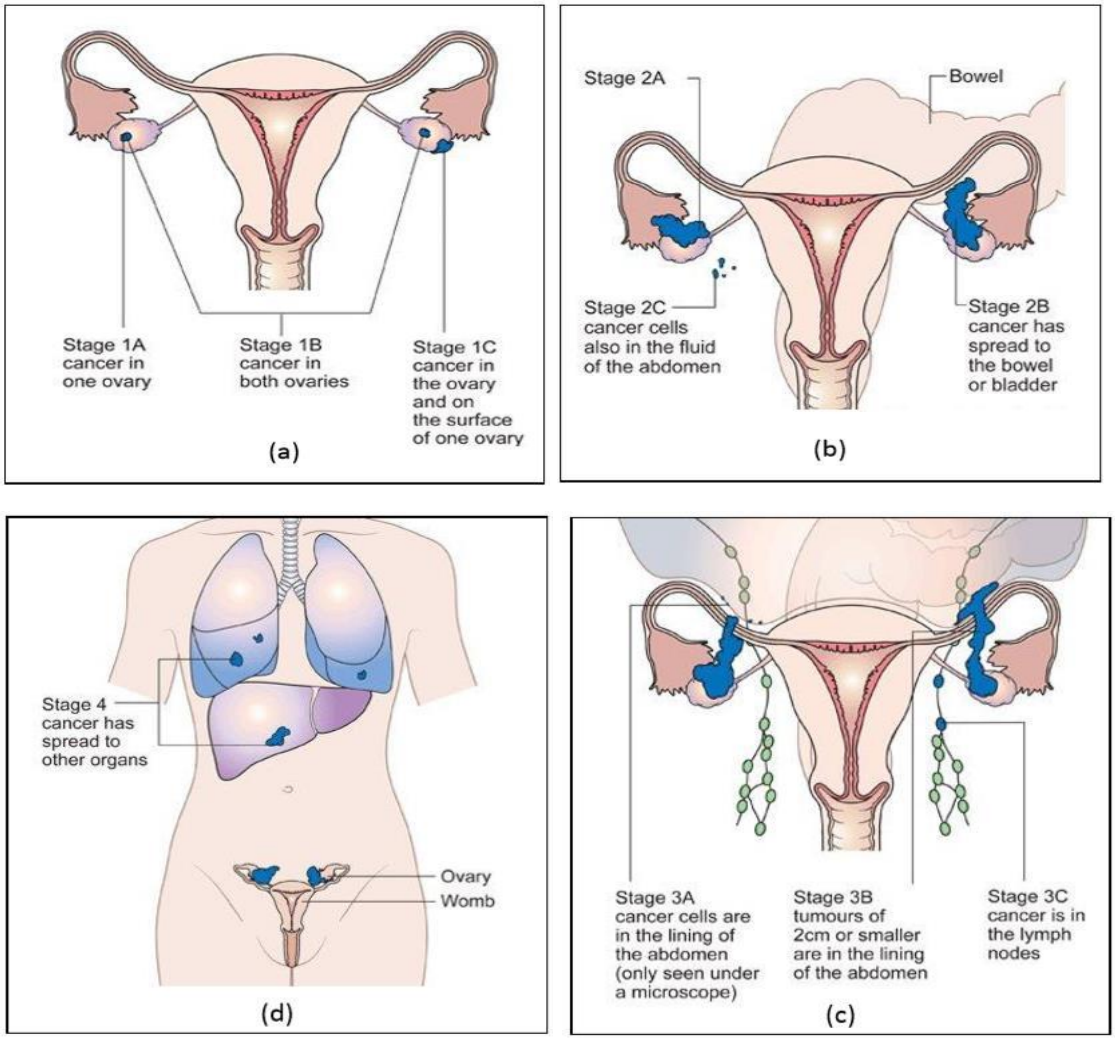


Figure 2- 8: Illustration of the extent of the spread of ovarian cancer - (a) - (d) for stages 1 to 4 (UK Cancer Research 2018).

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 at the ovaries.

2.2.4 Diagnosing Ovarian Tumours - Overview

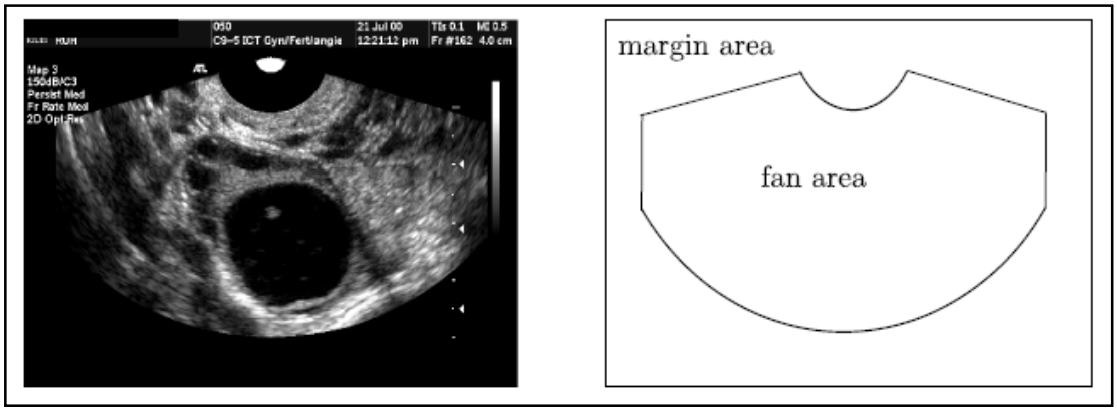
The established medical approach to diagnosing Ovarian cancer is based on several different types of tests, including a Physical examination, blood test (CA-125 level), biopsy, image testing, i.e. (ultrasound, CT scan, MRI), and so on. However, none of the tests on their own provide proof of presence or absence of ovarian malignancy. Providing appropriate treatment or relief depends on the accuracy and reliability of the diagnostic decision. The clinician also needs to have access to a reliable radiology assessment of the ovarian ultrasound scan, together with the blood test (CA-125). The timely availability of these assessments cannot be emphasised enough. The CA-125 is a protein in the blood that shows the level of CA-125.

As a tumour marker, this test is also useful in assessing the success of treatment because in some cases reduced the level of CA-125 is an indicator of successful treatment of cancer (American Cancer Society 2018). Nonetheless, the level of the CA-125 is not always a reliable indicator because not every woman who has ovarian cancer have a high level of CA-125 in her blood. Therefore, clinicians might consider doing further tests, such as ultrasound test to look at the size of ovaries, the normality/abnormality of the ovaries texture and if there are some cysts in the ovaries. All of these signs are very important to discriminant between different types of tumour (American Cancer Society 2018).

2.2.4.1 Role of Ultrasound Imaging in Ovarian Cancer Diagnosis

Ultrasound is presently one of the greatest widely used imaging modalities in medicine and has been used in medical imaging for over half a century (Hangiandreou 2003). It is usually considered as the ideal imaging modality for observation of the female pelvis (Quaia 2005). Technology developments have led to wide acceptance and use of ultrasound imaging. For example, the introduction of the transvaginal transducer has improved the diagnosis of ovarian tumours (Twickler and Moschos 2010). Analysis of ultrasound images can help to distinguish between benign and malignant lesions (Hamid 2011).

Each ultrasound image can be divided into two areas: a fan-shaped area and a marginal area. The fan-shaped area illustrations (see Figure 2-9) the image of interest obtained by ultrasound scan, but the marginal area contains other information, for example, the patient's name, the date that the image was taken and the ultrasound machine setting. (Harrington 2007)



(a) (b)

Figure 2- 9: (a) Ultrasound Image, (b) Fan Area and a Margin Area.

Ultrasound Images have some characteristics that are useful to diagnose ovarian tumours and to differentiate between benign, borderline and malignant tumours. The analysis of US images contains two parts. The first concerns are identifying morphologic features based on B-mode images like unilocular cysts, multilocular cysts, solid tumours, fluid, papillary projections, internal wall structure and acoustic shadows. The second part is based on Doppler images that provide information about blood flow. These indicators are useful for doctors to determine the severity of the mass or cyst. An advantage of using medical ultrasounds for identifying various diseases is that it is non-invasive, cost-effective, and the image quality is continuously being improved. However, there are still limitations with the ultrasound method. For example, background tissues and body fat can add noise to the ultrasound image, making it difficult to focus on the tissues or organs of interest (Hiremath, Akkasaligar and Badiger 2013)**.** Many sonographers utilise an automatic image analysis because it allows for more accurate images that can help with a diagnosis.

2.3 Design and Development of Automatic Medical Imaging-based Diagnostic Systems

Analysing medical images for diagnostic purposes is a typical example of pattern recognition and classification in computer vision. Over the years a variety of models for pattern recognition and classification have been developed in many different types of computer vision applications, all having a common multi-steps architectural design, but the steps are application domain dependent. Figure 2-10 below depicts the most common steps of such systems.

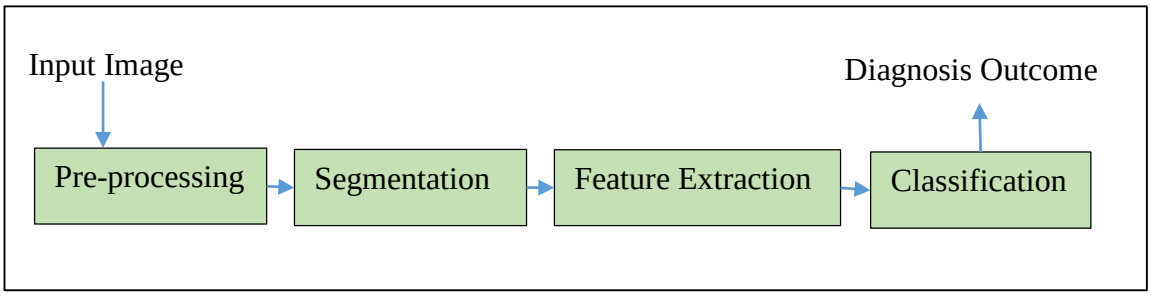


Figure 2- 10: Block diagram of the major steps of CAD.

As shown in Figure 2-10, a typical CAD system consists of three or four major processing steps. The pre-processing step suppresses or removes random noises from the input image that was included during the data acquisition and enhances the image by highlighting the image details to enable extraction of useful features from the image at a later time. This step often requires the use of suitable and sophisticated image processing functions to

achieve the set objective. After pre-processing, some form of image segmentation is normally performed to take out the relevant area of the image, known as the region of interest (ROI). This is because not all parts of a given input image are of interest; keeping the irrelevant parts in an image can affect the performance of the system at some later step. This step also requires sophisticated solutions based on computer vision and machine learning techniques because of the technical challenges encountered when the border of such a ROI is difficult to identify. The feature extraction step processes the segmented ROI image to extract quantitative descriptive data details that reflect the characteristics of the organ that the ROI depicts. The technical challenges at this step are to find useful features whilst at the same time limiting their number as appropriate. The final step in a CAD system is to use the features extracted from the various training images to build an effective model of classification for separating normal cases from medical anomalies of various stages and types. This step requires the use of machine learning techniques, particularly suitable supervised learning techniques, in building such an effective classification model. With machine intelligent solutions, a typical CAD system aims to address two major issues: *observer limitations* in relation to constrained human visual perception, fatigue or distraction, and limited knowledge and experience, and the *complexity of the clinical cases* where structures of medical anomalies overlap with structures of healthy cases (Lemaitre, et al. 2015).

2.3.1 Image Pre-processing

In many applications, the process of image data acquisition, the nature of the sensing devices (cameras) and the recording conditions often result in different quality images in terms of contrast, sharpness, and type of noise. Acquired images may also be of different sizes and/or resolutions. All these factors naturally expected to influence the analysis task to different extents, and the image pre-processing step is meant to counter/minimise the effect of these factors. Variation in image size is normally dealt with at the outset through existing software image re-sizing procedures while variation in the spatial resolution is dealt with by down-sampling or up-sampling using interpolations such as Bi-Cubic interpolation or even more sophisticated super-resolution techniques which are additionally useful in improving image quality (Al-Hassan 2014). In general, image input to pattern recognition schemes is assumed to be normalised in terms of size and resolution. Image quality enhancement is, generally, dealt with by domain-related procedures that require understanding the nature of the image degradation such as the type, characterises and source of the noise.

During ultrasound data acquisition, a certain type of background noise, known as Speckle noise, is widely known to corrupt the final image and the need arises for a noise reduction step. The primary objective of noise reduction in ultrasound imaging is to enhance the contrast between different tissues and organs. This step is meant to remove as many artefacts as possible while improving certain features of the image. This pre-processing step is also used in the manual diagnostic by physicians enabling them to understand and analyse the image better, for an efficient and more accurate diagnosis.

De-noising is particularly beneficial for distinguishing between different tissue types and is essential for the next stages of segmentation (i.e. isolating the target tissue/organ from other tissues) and features extraction. In the manual procedures, speckle noise reduction helps to detect the region of interest (i.e. the tumour) by the sonographer more reliably (Tamilkudimagal and Kalpana 2011). We shall now briefly describe the nature and causes of speckle noise in Ultrasound images

2.3.1.2 Speckle Noise and its Pattern and Physical Properties

The phenomenon of speckle-noise has long been an issue in ultrasound imaging since the 1970s (Abrahim, et al. 2012). Speckle noise is caused by interference of energy from randomly distributed scattering ultrasound waves dispersed from different locations and is generally observable as light and dark spots on an image. It differs from electrical noise because it is not arbitrary, despite its unusual appearance. Examining the object with a different transducer aperture, pulse length, or transducer angulation causes the speckled design to alter (Burckhardt, 1978; Goodman, 1976).

Figure 2-11 depicts the most-utilised model that elucidates the changes that occur upon intonation of a tissue. The tissue is a sound-absorbing medium that has caterers sourced from uniform elements and structures similar to or smaller than the ultrasound wavelength, such as tissue parenchyma (Thijssen 1993). They function by scattering sound waves within the tissue. Furthermore, scattering or speckling occurs when tissue particles that are comparatively small against the wavelength (i.e. blood cells) or have a different impedance are nearby. Additionally, ultrasound tissue absorption is an element attributed to pulse energy loss in scattering and refraction. Such loss arises due to attenuation, which is elevated due to depth and frequency and encompasses absorption, reflection, and scattering. Ultrasound with a higher frequency causes improved absorption and results in diminished visualisation.

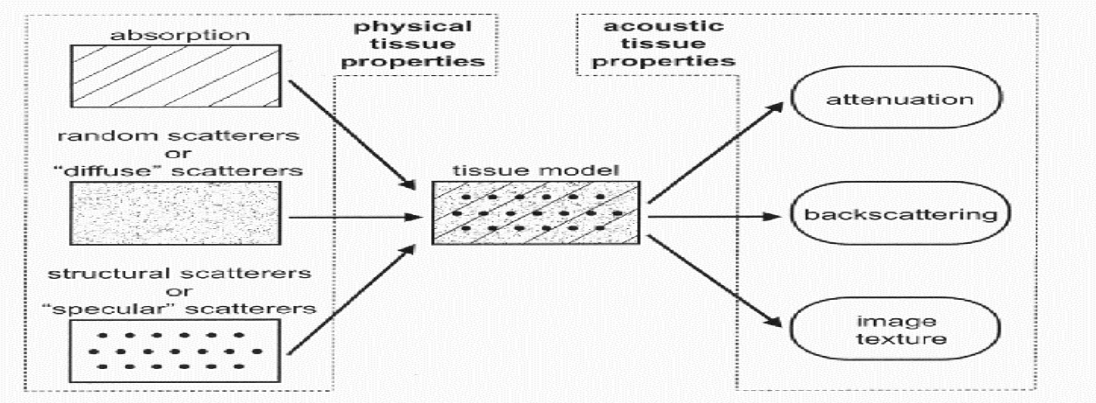


Figure 2- 11: The usual tissue model in ultrasound imaging.

Moreover, ultrasound propagation through tissue is influenced by other characteristics of the tissue itself, such as propagation speed, attenuation, and backscattering. The relaxation of the biological molecules generally causes ultrasound absorption, thereby converting mechanical energy to heat. Meanwhile, attenuation can also originate from scattering itself, which is the omnidirectional reflection of small-sized invariability in the tissue. Therefore, it is an outcome of absorption and scattering, as seen in Figure 2-11, which shows increasing attenuation in parallel to frequency, highlights the reliance on frequency.

2.3.2 Segmentation of US Images of Ovarian Tumour

One of the most challenging tasks in digital image processing is the image segmentation stage (Khiyal, Khan and Bibi 2009). The process of image segmentation targets a specific region of interest, a particular tissue, or organ that is deemed necessary for diagnosis. The area of interest can be isolated manually or automatically before image analysis. Automatic image segmentation relies on the discovery of digital features (in the spatial and/or frequency domain) that exhibit different quantitative values/statistics in the RoI than outside it. Therefore, in natural images segmenting objects/RoIs is done relies on discovering thresholds to separate the objects from the so-called background. This is rather challenging for US images of the ovary for a variety of reasons, and a close look at samples of such scan images demonstrate the futility of using the traditional thresholding approach. In chapter 3, we shall nevertheless describe and review some existing automatic and semi-automatic segmentation schemes specifically proposed for US ovarian scan images.

2.3.3 Feature Representation of Images

After segmenting the Region Of Interest (ROI) in a medical image, significant features need to be extracted. Feature extraction is another important step in enhancing an

ultrasound image and is based on pattern recognition. The primary objective of feature extraction is to identify specific information from the original image and characterise this information using a lower dimension. Redundancy in data can occur when the amount of data is too numerous to be properly processed. In this case, the input data can be decreased and represented by a set of features. If feature extraction is done correctly, the features set will contain the important information and the entire original data set will not be required. This method, based on pattern recognition, continues to grow in the field of ultrasound imaging (Krishnamachari & Chellappa, 1997; Kumar & Bhatia, 2014).

2.3.3.1 Texture-Based Image Features and Analysis

The texture of an ultrasound image can be defined as the smoothness and structure of the objects. Texture in an image is critical as it is a primary characteristic of the image that helps distinguish different image objects and is therefore used for the analysis and interpretation of images in general biomedical images and ultrasound images in particular (Kurani, et al. 2004). For the purposes of the current study, the texture is defined in terms of the spatial distribution and variation of the pixel grey value (intensity) of B-mode images. Various techniques can be utilised to identify a variety of texture features.

Texture analysis relies on efficient computation of the texture features and determining their properties of the components in the image (Mathias, Tofts, & Losseff, 1999; Nailon, 2010), as well as to define critical textural characteristics for an ROI. Texture analysis is typically used in the medical field to help distinguish between normal, healthy regions, and regions that require treatment. It can also be used to differentiate between various tissues and organs (Castellano, et al. 2004).

Texture analysis approaches are divided into the following categories: (1) Structural, (2) Statistical, (3) Model-based, and (4) transform methods. The structural approach was described by (Haralick, 1979; Levine, 1985) and is based on textures being represented by microtextures and macrotextures. Microtextures refer to well-defined texture primitives such as sharp corners and end of lines, whereas macrotextures are defined in terms of a gradient of spatial arrangements of the microtextures. In this method, the texture is described based on the primitives and the rules for placement. The main advantage of this approach is that it offers a sufficient representation of the ultrasound image. For this reason, this method is better suited for synthesis tasks, rather than analysis tasks. For analysis of bones where fine details in the bone microstructure need to be identified, a mathematical morphology analysis may be used (Chen & Dougherty, 1994; Serra, 1983).

In addition to the structural approach, a statistical approach can be used. While the structural approach analyses the texture of an image using a hierarchy, the statistical approach focuses on non-deterministic properties of the various levels of grey in the ultrasound image. This method is considered to have a higher level of discrimination compared to other methods, such as the transform-based method and the structural method. Statistical methods to analyse texture have also been used by (Julesz 1975) to identify texture within human tissues.

Gray level images can be distinguished according to their texture, but only if different in the second-order moment. Automatic processing may be performed for second-order moments, but third-order moments require careful consideration. Texture analysis features commonly found at the second-order may be an obtained from a co-occurrence matrix, which has been proven effective in discriminating among textures found within ultrasound imagery. In fact, multidimensional cooccurrence matrices were demonstrated to outperform wavelet packet, which is a transformation based technique, and applied to texture classification schemes (Hassner & Sklansky, 1980; Haralick, 1979; Lerski, et al., 1993; Valkealahti & Oja, 1998).

The model-based texture analysis is used by a number of researchers (Chellappa & Chatterjee, 1985; Derin & Elliott, 1987; Manjunath & Chellappa, 1991). This method uses fractal and stochastic models that can help to understand the texture, and this is of particular importance for ultrasound image analysis. Image analysis can occur based on an estimation of the parameters of the model. However, estimation of the stochastic model parameters often requires complex computations, which can be difficult to achieve. The fractal model has proven to be helpful for analysing natural textures and is often used for the analysis and identification of textures (Pentland, 1984; Kaplan & Kuo, 1995; Chaudhuri & Sarkar, 1995). A limitation of the fractal model is that it cannot discriminate local image structures, and the orientation cannot be selected.

Textures can also be analysed with transform methods. The most notable transform methods include Fourier (Rosenfeld & Weszka, 1976), Gabor (Daugman, 1985; Bovik, Clark, & Geisler, 1990), and the wavelet transform method (Mallat, 1989; Lerski, et al., 1993; Lu, Chung, & Chen, 1997). These methods represent the image using a coordinate system that has been interpreted to be fairly similar to specific textural properties, such as size or frequency. The Fourier transform method is deficient in spatial localization and is not commonly used due to this failing. In contrast, the Gabor filter performs better in spatial localization compared to the Fourier transform, but its downfall is that it uses a

single filter resolution, which's hard to apply to natural textures. Of the transform methods, the wavelet transform method offers the most benefits. The wavelet transforms feature is more versatile to the other methods because the spatial resolution can be adjusted until the optimum scale is found. This allows for a better representation of textures. In addition, the wavelet function offers a variety of choices so that texture analysis can be used for any type of application. Due to this, the wavelet transform method may be the best method for segmenting textures. One major limitation of this method; however, as (Materka, Strzelecki, & others, 1998; Lu, Chung, & Chen, 1997) noted is that the wavelet transform is not translation invariant. However, various techniques can be utilised to identify texture features, which will be described in detail in next chapter.

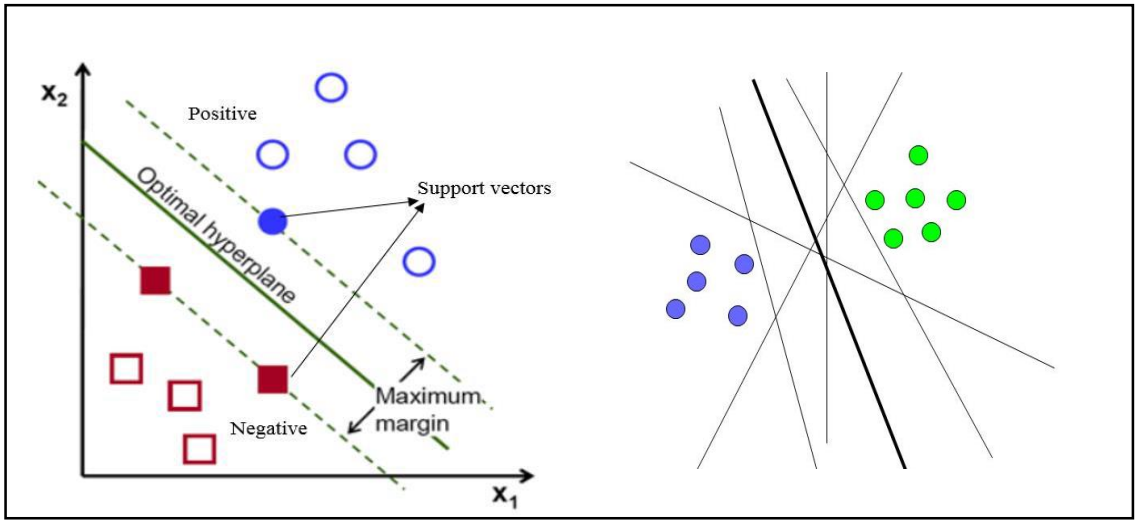
2.4 Classification Techniques

Classification is the process which used to distinguish between different types of classes via labelling for each similar data set with a certain label to discriminate it from other classes. There are two stages in image classification, training and testing. The training stage is used a set of sample images with identified class labels to train the classifier for instant Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), decision tree (DT), etc. However, after the classifier is trained, the testing stage is used to predict the image class. Therefore, the classification results are dependent on the prediction class of image; thus, of the predicted class is similar to the known class of the test image, the classification is correct. Else, it is not correct. In this thesis, we have used the most two popular classifiers, which are Support Vector Machine (SVM), k-Nearest Neighbour (k-NN). The usage of them in many areas of research, for instance, image classification, text, medical image recognition, face recognition, etc., make them popular classifiers. (Cristianini and Shawe-Taylor 2000)

2.4.1 Support Vector Machine (SVM)

The SVM method was first proposed in 1995 by Vladimir N. Vapnik and Cortes, with the algorithm later created by Vapnik (Cortes and Vapnik 1995). The basis of this method is that decision plane, which categorises objects into a positive and negative class, defines decision boundaries. Figure (2-12a) displays an example of the decision plane, where a boundary is shown between the red colour and blue colour. These two colours symbolise two classes. If any object falls on the left side, it will be classified as RED. Similarly, any object that falls on the right side is classified as BLUE. As the white circle in the figure demonstrates, because it falls on the right side, it will be considered BLUE. The line that separates the two classes is referred to as a hyperplane.

Many different types of hyperplane can separate two classes. The optimal separating hyperplane, as proposed by (Han, Kamber and Pei 2011), is the hyperplane that maximises the distance between the plane and each class' closest data point, thus, maximising the separation of the classes. In Figure (2-12b), the optimal separating hyperplane is shown as a black line. The features along the boundaries are called "support vectors". It is the support vectors that are used to differentiate the two classes.



(a) (b)

Figure 2- 12: Diagram Representation of the Principle of SVM. (a) SVM attempts to Maximise the Margin from the Hyperplane to find the Best Separate between two Classes (red positives from blue negatives) (b) Optimal separating Hyperplane.

There are different steps to classify a new case: -

i. Two different subsets will be used, which are training and testing. The training was spilt form the dataset regarding class label to train the SVM and also to find the best hyperplane based on margin maximisation among two classes.

ii. All training points consistent with the mean (μ) and the standard deviation (σ) was standardized to find the support vectors. This process is done on all training samples for each feature component domain:

$$Si = xi - \mu / \sigma \tag{2.1}$$

iii. The score of the test case will be calculated based on the Lagrangian formulation

$$Sc(x) = \sum_{i=1}^{n} \alpha i \; k(si, x) + b \tag{2.2}$$

Where *Si* is a support vector, *αi* is the weight of *Si*, *n* is the number of support vectors, *k* is a kernel function. In the case of a linear classifier, the kernel function *k* is simply the inner product <*si*, *x*> and the bias b.

i. If the score is less than zero, then *x* is classified as positive. Otherwise, it is classified as negative.

2.4.2 k-Nearest Neighbour (kNN) Classifier

The k-Nearest Neighbour (kNN) classifier method was first proposed by (Fix and Hodges Jr 1951)and is still being used today for pattern recognition. This works by classifying an unknown feature to the category in which the highest percentage of its neighbours belongs. In this first step, patterns that are used for training, also referred to as feature vectors, are oriented in the feature space that is multi-dimensional. The distance is calculated between the feature vector and the unknown feature using a distance metric. The most frequently used distance metric is the Euclidean distance metric (Webb 2003). The kNN algorithm for classification is outlined as follows:

1. Compute the distance between test case X = (x1, x2, …, xn) and each training templateY = (y1, y2, …, yn) using a distance measure such as the Euclidean distance D2:

$$D_2\,(X, Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}^2 \tag{2.3}$$

2. Sort all distances in ascending order and identify *k* nearest neighbours of *X: Y1, Y2, …, Yk*, where *k > 1*

3. Assign X the class label of the majority among *Y1, Y2… Yk,*

In Figure 2-13, an example of the kNN algorithm is illustrated. If K is equivalent to 2, the green dot (unknown vector) will be categorised with the red triangles. However, if K is equivalent to 3, it will be categorised with the blue squares. Note that when K=5, the blue squares win, as the majority of the neighbours are also blue squares (i.e. 3 nearest blue squares, only 2 nearest red triangles.

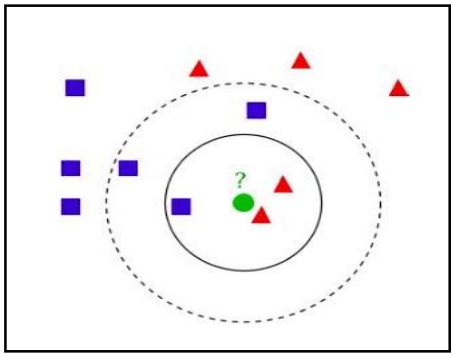


Figure 2- 13: Illustration of the working of kNN classifier.

2.5 Summary

This chapter reviewed the background on medical imaging and computational besides 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 malignancies in women and is the leading reason for death from gynecologic tumours. 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, maybe 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 might 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 an early stage. CAD could 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 shall review existing work related to this thesis objectives.

Chapter 3

Computerised Ovarian Tumor Diagnostic – A Literature Review

In chapter 2, we gave a brief description of the different types of ovarian tumours. We also highlighted the crucial role of B-mode ultrasound images within the standard clinical approach to ovarian cancer diagnostic. Reliable interpretation and analysis of ultrasound images are dependent on the medical expertise and experience of the observer, which may be subject to considerable variation. Therefore, the manual process involves multiple subjective decisions that are subject to inter- and intra-observer differences which may lead to difficulties and even errors in the diagnosis stage, (see (Wang, et al., 2002; Giger, Chan, & Boone, 2008)). These concerns are compounded by the increased demand for rapid analysis of Ultrasound images necessitated by the desperate need for early detection. Shortage of highly experienced clinical experts can only be mitigated by developing reliable computerised ultrasound image classification algorithms as diagnosis decision support tools. In chapter 2, we also described the design strategy for such systems and discussed briefly their main requirements and ingredients. This chapter is a follow up on the previous one by reviewing existing work in relation to the various stages of automated diagnostic tools pipeline. We shall highlight the shortcomings/limitations if any, that this thesis aims to overcome and point out our approach to dealing with each stage. We first review in section 3.1 the literature relevant to the main challenges and solutions associated with pre-processing, and then discuss existing work in section 3.3 on tumour ROI segmentation. Most relevant existing feature representations of ROI from ovarian scan images will be reviewed and briefly described in section 3.4, prior to reviewing existing automated or semi-automated diagnostic decision systems for ovarian ultrasound images which will be explained in section 3.5.

3.1. Ultrasound image Pre-processing

Algorithms for the analysis of Ultrasound ovarian images must have consistent performance regardless of the source of the image data which might be generated by different devices operated by different persons who may have a different level of training. The process of image acquisition and transmission of the scanned images may result in variation in image quality that may have an adverse impact on the feature extraction step

and subsequently on the analysis of these images. Accordingly, some pre-processing need to be applied to the scanned images to ensure compatibility with the system requirement and reduce the adverse impact on analysis.

Image size and formats need to adhere to a strict rule, and hence, reliable image re-sizing procedures are essential. The more serious issue that needs to be dealt with is that of enhancing image quality to a standard level range. The most concerning aspects of image quality is that of the presence of noise, artefact and other forms of image degradation. Noise and artefacts are generated in different types of medical image modalities, and in the previous chapter we pointed out that the main concern in this respect for US images is that of the presence of speckle noise.

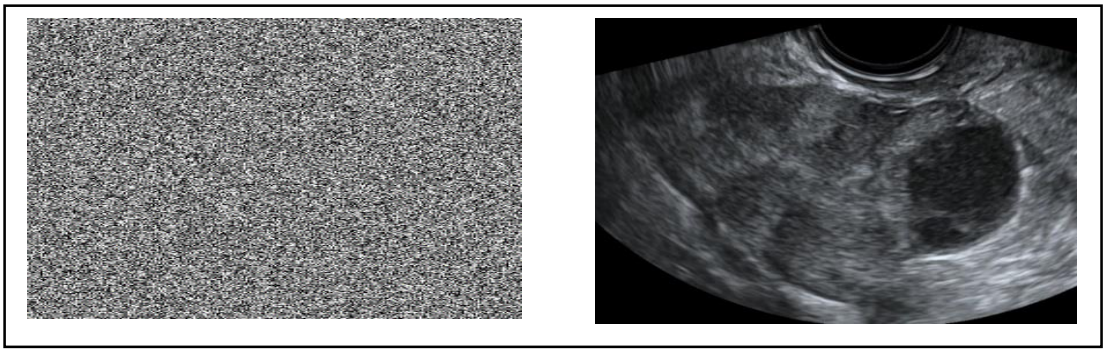
3.1.1 Image Noise - Nature characteristics

Identifying the nature of the noise is vital when selecting the type of filter to use when rectifying an image. Noise is visible in the image through unwanted variation observed in brightness or colour. One example is electronic noise, which can be caused by the sensors of a digital camera or a scanners' circuitry. Noise can appear because of a slow shutter speed or from light generated through high exposures (Hedrick and Hykes 1989). There are other types of noise, such as salt & paper, quantisation noise, and impulse noise. This study focuses on multiplicative noise, also known as speckle noise.

Speckle is a type of locally correlated multiplicative noise that distorts ultrasound images and makes observations somewhat challenging. The problem has been around from the early days in the 1970's and has been the subject of many research efforts on how to deal with it. Its adverse effect is more apparent when dealing with small low-dissimilarity lesions. It is important to remove speckle from images while preserving vital information. Speckle noise reduction methods can enhance the image and allow for more accurate manual diagnosis of medical conditions (Tamilkudimagal and Kalpana 2011).

Medical ultrasound images rely on biological tissues scattering or reflecting incident sound and the concepts of scattering and reflection are linked. If a particle reading is below a sound's wavelength, it is identified as scattering, while when it is larger than the wavelength, it is known as reflection. This variation, like the acoustic waves, is caused by differences in the medium's thickness or compressibility. Once the back-scattered acoustic pulses arrive back to the transducer, they receive acoustic energy with constructive and destructive involvement, which causes a granular pattern referred to speckle noise (Dangeti, 2003; Burckhardt, 1978). It has a negative impact on image quality because it hides and distorts essential features, especially around the edges.

Distortion makes image segmentation and post-processing operations more difficult and can reduce the diagnostic value of the image (Zhu, Ni, Li, & Gu, 2009; Xie, Jiang, Tsui, & Heng, 2006). Figure 3-1 shows the effects of speckle on ultrasound images of an ovarian tumour. The reduction of speckle noise is vital for effective automatic processing and analysis of ultrasound images.



(A) (B)

Figure 3- 1: Illustration of speckle noise and its effect: (A) Example of speckle noise and (B) US image of an ovarian tumour corrupted by speckle noise. (Dangeti 2003).

3.2 A Review of Image De-Noising Algorithms

Depending on the characteristics of the noise in an image, a variety of de-noising schemes have been deployed that generally adopt spatial and frequency domain filtering (Dangeti, 2003; Zhu, Ni, Li, & Gu, 2009). In (Nobi 2011) presents an efficient and straightforward method for noise reduction from magnetic resonance (MR) images and ultrasound images. The authors in (Nobi 2011) proposed a modified median filter with additional features and shown that their results outperform three commonly used image filtering algorithms based on order-statistics filters (i.e. mean, median and midpoint).

Wavelet transforms, and filters have been used for de-noising several types of biomedical images (N. a. Attlas 2014). These approaches have also been used for de-noising natural images. Generally, the work by suppressing certain high-frequency wavelet coefficients that isolated from edge coefficients that tend to form spatial clusters in the high-frequency sub-bands. Having removed such isolated coefficients, the de-noised images are obtained by inverting the wavelet transform using the modified sub-bands. Normally, the detailed Low-Low sub-band will not be modified.

Another approach for de-noising, known as the non-local means scheme, has been recommended for use in reducing speckle noise in ovarian ultrasound images (Khazendar, et al. 2015). The non-local means scheme assumes that the presence of many intensity similar patches throughout an image are indicators of the effect of noise on the centre of

these patches and therefore reducing variation in the intensity of those central pixels helps reducing noise. It works by replacing individual pixels in the image by weighting the average of all pixels with comparable neighbourhoods. The benefit of using this filter is that all-important information is retained while speckle noise is reduced. This filter also takes advantage of the properties of the image, such as redundancy and self-similarity. The non-local means filtered image is further enhanced with a negative transformation that computes the absolute difference between the improved image and the original image. This results in clearer texture and improved quality, especially around the edges and important features. For Ultrasound Ovarian scans, this scheme seems to provide improved details within ovarian cysts.

The Wiener filter is a well-known inverse transform that is used to estimate and remove unknown image degradation models, including noise. In (Garg, et al. 2011) the Wiener filter was used to remove speckle noise by thresholding the discrete wavelet transform domain. First, the multiplicative noise was converted to additive noise by applying a log transform. Next, the image is Wiener filtered, the output of which is subtracted from the log-transformed observation.

A comparative study was conducted, in (Attlas and Gupta 2014), using different types of de-noising filters for ultrasound images, including Weiner filter, Bayes wavelet filtering and Morphological filtering technique. Different statistical quantity parameters measurements were used to assess enhanced images. It was found that the Morphological filter outperformed the alternatives. Initially, the different filtering methods were compared with the de-noising schemes that combined discrete wavelet transform (DWT) with the Wiener filter and (DWT) with median filter, as well as, one combination of both Wiener filter and the median filter used together to remove the speckle noise. The authors in (Attlas and Gupta 2014) again compared the results using statistical measurements and found that the DWT performed better than the alternatives.

In another study, a Wiener filter was combined using a wavelet domain with soft thresholding as a comprehensive technique (Sivakumar 2010). The authors in (Sivakumar 2010) compared the efficiency of the wavelet-based techniques for de-speckling with other classical speckle reduction filters. Statistical quality measures were also used to determine the performance of these filters. The results were obtained by displaying the various filtered US images according to their statistical measures as representing the visual quality of the images. The Wiener filtering with Bayes Shrink thresholding technique in the wavelet- domain performed the best out of filtering techniques.

Another study, a hybrid filter was proposed by (Metha 2017) based on wavelet transforms and winner filter to reduce the effectiveness of the speckle noise in US images. The multiplicative noise was converted to additive noise by applying a log transform. After that, the wavelet was used to decomposition the images into four levels. The statistical parameters and threshold value were computed for each sub-band. Then, the soft thresholding used on all subbands and then inversed the WT to produce the de-noised image. The Weiner filter applied to the results of the WT images for the final de-noised image.

In another study, different types of speckle-noise reduction methods were compared based on statistical quality measures to de-noised the US images of polycystic ovary. The authors in (G.Vasavi and S.Jyothi 2019) compared between median, Wiener, Gaussian, anisotropic diffusion, Non-Local Means (NLM) and others to find the best filter to remove the speckle noise. After evaluating all noise removal filters, the authors found that the NLM filter is outperformed the reset methods.

Despite limited success, a recent survey conducted by (Zhu 2009) summarised and highlighted some significant disadvantages of existing de-noising research. Firstly, each filter is sensitive to a specific model of noise. Secondly, most filters succeed in reducing noise in smooth and background areas, but they do not perform as well when enhancing features like edges. Thirdly, most existing filters are not adaptive and have the "smoothing effect" on borders and other distinct image features. Therefore, noise removal for medical ultrasound images is still a challenging task as it may have undesirable effects on image texture, which in turn may have an unexpected effect of diagnostics.

In this thesis, we shall show experimentally that diagnostics decision is tolerant to a certain level of speckle-noise, and we shall develop an adaptive speckle de-noising approach that is only applied to parts or all image whenever needed. We shall demonstrate that this is an efficient scheme that inhibits sufficient amount of speckle noise in ultrasound images of ovarian tumours without adversely influencing the image texture.

3.3 Segmentation of US Images of Ovarian Tumour

One of the most challenging tasks in digital image analysis is that of segmentation stage (Khiyal, Khan and Bibi 2009). The process of image segmentation aims to isolate specific region(s) of interest (e.g. a particular tissue or organ) that is deemed necessary for diagnosis. Accurate segmenting ovarian tumour regions in ultrasound scan images is

expected to enable us to extract the most relevant digital texture features that discriminate different types/stages of tumour.

The area of interest can be isolated manually, but success is dependent on the expertise of the specialist operator. In general, automatic image segmentation relies on the discovery of digital features (in the spatial and/or frequency domain) that exhibit different quantitative values/statistics in the RoI than outside it. In natural images, visually objects of interest are easy to distinguish from their surrounding regions in terms of colour, texture and semantics. Therefore, segmenting objects/RoIs in many such cases relies on discovering thresholds that separate the objects from the so-called background. The main shortcomings of these approaches to segmenting natural images relate to variation in image quality and recording conditions, the presence of noise and occlusions, among other factors. In biomedical images and ultrasound ovarian scan images, the challenge of segmenting areas of abnormalities is rather much tougher than the case of segmenting natural images for a variety of reasons. Indeed, automatic ovarian tumour segmentation still requires further investigation, with (Sohail, Rahman, et al, 2010; Hamid, 2011; Khazendar, et al., 2014; Khazendar, et al., 2015; Aramendia-Vidaurreta, et al, 2016) reporting the subject as very challenging.

Several approaches were nevertheless investigated and proposed for segmenting the ROI in ultrasound images in relation to detecting different kinds of ovarian abnormalities. In (Harrington 2007), the authors proposed a method that can automatically segment ovarian follicles. Although, ovarian follicles generally appear dark on an ultrasound image but not every dark connected region is a follicle. Instead of thresholding, the authors used Harrington's geometric active contour models to isolate the dark regions in the ultrasound. Following this, to determine if the dark region is an ovarian follicle, a Bayes classifier method is used. Although this maybe a plausible method to segments cysts septum, unfortunately, the appearance of non-solid visible tissue inside a large dark area may not represent a common border between two chambers.

Finally, we observe the need to extend our review to other ROI segmentation in Ultrasound images of non-ovarian tissue in relation to diagnostics of cancer in other body tissues/organs.

In a study on segmenting prostate regions, (Rafiee, Salimi and Roosta 2008) suggested a similar semi-automatic US image segmentation method for automatically isolating the prostate boundary. Specific seed points are chosen from the prostate, followed by an active contour procedure. Numerous researchers have validated this technique and have

yielded similar results to a manual segmentation method conducted by experts.

The left ventricle of the heart has furthermore been segmented using a method recommended by (Landgren, Overgaard and Heyden 2013). This semi-automated technique is based on the snake method, where two set points are identified, and segmentation occurs over a whole cardiac cycle. In an alternative study (Deshpande, et al. 2013) Offered another automatic segmentation method for two-dimensional foot ultrasounds. This method is completed in three steps. In the first phase, the image is enhanced through an anisotropic diffusion filter and the improvement of contrast. The second step involves segmentation through an active contour technique, yielding a binary image. In the final step, boundaries are enhanced, and undesirable regions or objects are taken out of the image.

In (Saranya 2016) an automatic follicle detection system is proposed using an Adaptive Particle Swarm Optimization technique that overcomes the challenges in detecting follicles from the ultrasound ovary images. In this paper, an idea is introduced to optimize the objective function described by the modified Otsu method using the Particle Swarm Optimization (PSO) and proposed Adaptive Particle Swarm Optimization (APSO) approaches. The problem of thresholding is reduced to an optimization difficulty in order to search for the thresholds that maximize the between-class variance. This approach is when finding an optimal set of thresholds with a larger between-class variance than the other approaches. This method was applied to 158 ovarian images. They discussed the result in two ways: 1) measure some geometry feature and then compare the proposed model measurements with both expert measurements and other existing methods measurements. 2) Showing some segmented images. Based on the result, we can clearly see the effect of the proposed APSO method.

Although, the above research investigations help understand the challenges of segmenting ovarian tumour from Ultrasound images, but our objectives are likely to benefit more from research into other types of ovarian abnormalities. The use of Ultrasound scans of the ovary during pregnancy is by far the most frequent practice, and segmenting the Gestation Sack (GS) is a major step in ant research that aim to computerise miscarriage, in (Ibrahim, Al-Assam and Du, et al. 2016), developed an automatic segmentation of the GS from a static B-mode image scan of the ovary for early identification of miscarriage cases. The algorithm has been developed through a multi-stage refinement that exploits the well-established geometric features, shape and content of the GS. Recognising the adverse effect of speckle noise on these features, the scheme begins by suppressing the

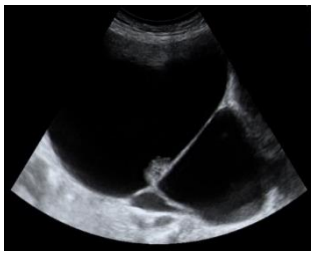
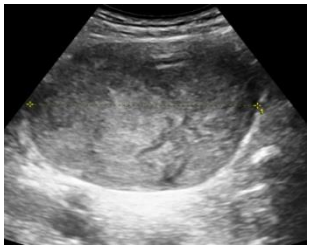
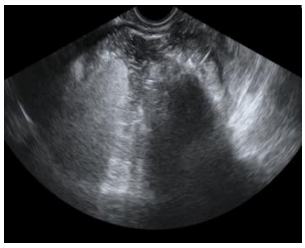
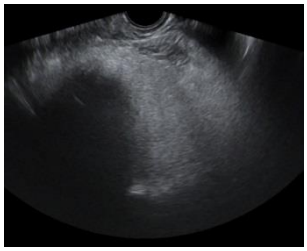
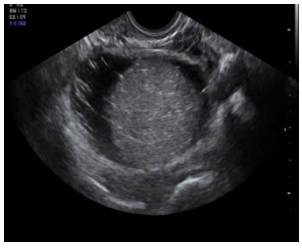
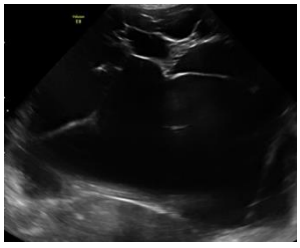
speckle noise and to enhance the image prior to applying wavelet transformation. Then, the mean value is applied as a threshold to binarise the image, followed by filtering unnecessary objects based on their circularity, size and mean of greyscale. The mean value of every object is afterward used to further select candidate objects from a number of potential candidate regions. Despite considerable success, the detected region in some cases have a missing part of the boundary or do not fit the expected shape of a GS. As a result, a *Region Growing* technique was used as post-processing to improve identification of the GS. Unlike the GS region, tumour cysts regions vary in shape, size and content from one case to another.

In (Ibrahim, Al-Assam and Jassim, et al. 2017), the author of the above scheme investigated the use of a multi-level trainable machine learning segmentation technique with the aim of reducing false positives. It was noted that the histogram of oriented gradients (HOG) in blocks away from the GS exhibit different characteristics to blocks on the border or inside the GS. Therefore, the first step of their proposed segmentation scheme is based on training a neural network classifier with blocks from outside GS and blocks from the border of the GS. Besides segmenting the GS, this scheme helps estimate the size of the GS. Interestingly, they also used the second level of trainable segmentation to detect and measure the size of the Yoke Sack (YS) inside the GS. The efficacy of the proposed solution was tested by examining automatic size measurements of GS and YS in comparison with measurements obtained from the gynaecologist. Results from testing 199 ultrasound images using the proposed solution show that it is effective in accurately measuring and identifying the correct stage of pregnancy.

We note that the trainable segmentation approach, adopted in (Ibrahim, Al-Assam and Jassim, et al. 2017) for the GS segmentation, together with the active contour shape approach may have the potential for success in relation to ovarian tumour region of interest. For that, we need to use one or more image texture features that can distinguish blocks of normal ovarian tissue from tumour (Benign or Malignant) blocks. However, the amount of work needed to develop such an automatic segmentation scheme is deemed to require somewhat extensive research that would be better taken after achieving success in showing the viability of using AI and machine learning for diagnostic purposes. We note that segmenting tumour masses is somewhat very different and more challenging than segmenting the GS and its content during pregnancy. In fact, the existing dataset for US ovarian tumour scan images displays wide disparity in tumour shape, size, and position, and it can be difficult to differentiate if the region of interest and background

region do not contrast (Zhang 2006; Ozgen 2011). Factors that impede and complicate automatic ovarian tumour segmentation includes: first, **the image quality:** it influences segmentation processes due to the unclear demarcation of the tumour border. Second, **the size of a tumour:** it may be observed to be large in the images of the dataset, and complicated by an unclear delineation of a tumour inside the fan area**. Third**, **the similarity of texture:** this may be present within tumour location and the neighbouring background in some of the images, leading to challenges in presenting an automatic demarcation of the tumour border, especially if there is only one static B-mode image**. Fourth, different textures may exist inside one tumour:** this may occur in some instances, whereby several elements having specific inner borders (e.g. multilocular cyst) may be present within one tumour**. Finally, the combination of the above:** a combination of all features may be observed in a single image. The above factors justify the adoption of a semi-automatic approach for tumour segmentation. Table 3-1, below, show samples that highlight some of the challenges. In this thesis, we shall not focus on this issue and opt for using manual cropping of the ROI by the medical operator.

Table 3- 1: Lists some examples that illustrate the different problems mentioned above.

The large size of a tumour	Poor quality of the image	Tumour and the background are the same texture
		
Undefined the border for a tumour.	The different textures inside a tumour.	Large ovarian tumour undefined the border.
		

The issues highlighted above are listed accordingly in Table 3-1. Despite such concerns, this work is adamant regarding the feasibility of automatic segmentation, but the scrutiny is directed more towards the reliability of the feature extraction processes. Therefore, the segmentation procedure was undertaken manually as per guidance from domain experts;

the resulting outcomes were then assessed and verified by the experts. ImageJ was utilised to perform manual segmentation due to its frequent use in biomedical image analysis. Additionally, Figure 3-2 illustrates instances of ovarian tumour images that were manually segmented, which are defined using yellow borders.

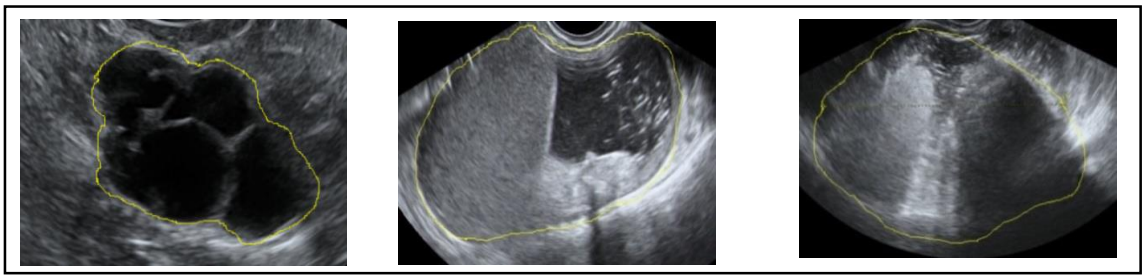


Figure 3- 2: Manually segmented ovarian tumour regions marked by yellow borders.

3.4 Texture Representation of Segmented Ultrasound Images

Once the region of interest (ROI) is segmented, manually or automatically, and confirmed validity preferably in consultation with medical experts the process of feature extraction is the next step to create a feature vector representation of the segmented region and preferably quantifying texture of the image in a form that helps discriminate Benign tumours from Malignant ones. In the medical domain, interesting results have been reported on using image texture analysis for diagnostic purposes. In ultrasound images, many studies focus on characterising B-Mode images.

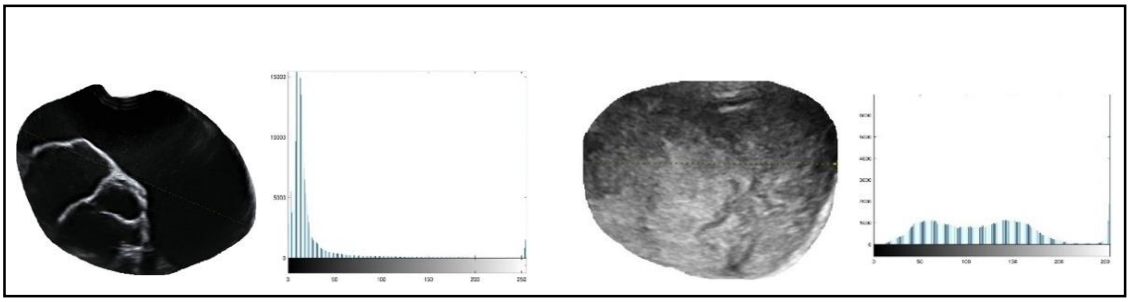
Texture analysis of ultrasound images is governed by the principle that if the disease process affects the structure/texture of tissues, then the diseased tissue should reflect ultrasound wave signals differently to healthy tissue (Morris 1988), i.e. texture features extracted from the US scan of diseased tissue are expected to differ from those obtained from the US scanned healthy/normal tissue. The transformation of cancerous tissue, for example, results in the changes in the tissue characteristics due to nature of cancerous cells such cells continue to divide, grow and create own networks of cellular structures necessary for their growth. Therefore, it is expected that the textural features of cancerous tissue and healthy, as well as abnormal benign, tissue, be noticeably different.

In chapter 2, we described the classification of different models of image texture descriptors. One such model was the structural-based texture divided into microtexture (e.g. sharp corners and end of lines primitives) and macrotextures defined in terms of a gradient of spatially arranged micro-texture primitives. Statistical texture models, on the other hand, focus on non-deterministic properties/moments of distributions of grey levels in the ultrasound image. Other image texture representation methods include Model-

based and transform-based methods. However, the statistical-based texture features have been more popular in analysing medical images for identification of texture within human tissues, (Julesz 1975). It is worth noting that these different models are not exclusive. For example, image structural and transform-based textures are often quantified through statistical analysis of the distributions of the obtained structural primitives and transformed coefficients. In the following subsections, we shall attempt to review several commonly investigated image texture features, but according to slightly modified categorisation, we highlight their use in representing abnormal regions of ovarian ultrasound images. In the next chapter, we identify a list of several such features extractable from the spatial domain of B-mode Ultrasound ovary scans, formally describe their computation, and test their performances in terms of distinguishing benign from malignant ovarian tumours.

3.4.1 Statistical based Image Texture descriptors

There are many image texture representations within this model. The simplicity and computational efficiency of statistical texture feature explain their popularity in general image analysis applications, including analysis of medical images (e.g. see (Khazendar, Shan 2016)). The simplest such statistical model is the histogram-based representation of image texture that consists of six parameters extracted from the image (or the RoI) spatial domain: mean, variance, skewness, kurtosis, energy and entropy. Together these features provide information about the shape and distribution of the image histogram, and it has been suggested that together, these measurements have a high discriminative power to distinguish between different images. Figure 3-3 shows examples of the histogram distributions for various ultrasound images.



(a) **(b)**

Figure 3- 3: Histograms of ultrasound ovarian tumour image (a) Benign, and (b) Malignant.

The simple Histogram texture descriptor suffers from the fact that a large number of different images can be constructed from a histogram. A more sophisticated statistical-based texture feature representation of images that reduces the effect of this problem is

the so-called Histogram moments, a popular example of which is the seven Hu moments, (details in next chapters). An alternative statistical-based texture descriptor is provided by the Gray Level Co-occurrence Matrix (GLCM). The GLCM is second-order statistical texture representation defined by the conditional joint probabilities of all pairwise combination of spatial domain intensities, and it is often calculated in image blocks. These statistically-based texture descriptors have been used (either singularly or jointly) for retrieval and classification of abnormalities in ovarian ultrasound images, (see, e.g. (Sohail, Rahman, et al, 2010; Ibrahim, Dheyaa Ahmed, 2018)).

For natural images, some or all these statistical moments are used for texture analysis and classifications in the spatial domain as well as frequency domain and many transform domains. In fact, histograms and probability distributions form a common representation of transformed data. Examples of transformed texture features that depend extensively on histogram analysis of non-spatial domain image representation include Gabor wavelet transform and Histogram of Oriented Gradient, as well as the texture model in the next subsection.

3.4.2 Transform-based Texture Descriptors

A variety of natural image transforms have been developed in the past as a complementary representation of image content and have been used for different image processing/analysis applications as well as security procedures. Such transforms include Frequency domain transforms (e.g. Fourier and wavelet), gradients and Local Bina Patterns transform (LBP). These types of feature vectors are commonly used in many texture-based image analysis applications, including for diagnosing certain types of ovarian tumours, (e.g. see (Khazendar, et al., 2014; Khazendar, et al., 2015)). The LBP transform/concept was introduced by (T. M. Ojala 1996) in relation to face detection and recognition. The LBP transform of an image acts as a non-linear filtered version of the image, whereby each pixel is replaced by a byte representing the order relation between the pixel value and its 8-immediate neighbours scanned in a clockwise manner starting from to top left corner. If the neighbour is $\geq$ than the centre pixel, then the corresponding bit is replaced with 1 otherwise it is replaced with 0. This is normally referred to as the (8, 1), form LBP image code the analysis of which is a transformed based texture model. This LBP texture model is often quantified by different histograms of the LBP codes but the simplest being the 256-bin histogram created for the ROI image. This allows for easier identification of texture information in the ovarian masses because the LBP captures the local changes in the grey values throughout the image.

A variant of this representation of transformed based texture is offered by a 59 bins histogram representing 58 bins for the so-called uniform LBP (bytes that have no 1's or a single run of 1's), and the 59th bin represents all other non-uniform LBP codes. Pixels that have ULBP codes indicate primitive texture features, e.g. corners, spots, and end of lines/edges in digital images. Figure 3-4, below, illustrates two cropped images of ultrasound scans of ovarian tumours together with their LBP transformed images as well as their 256-bins and 59-bins histograms.

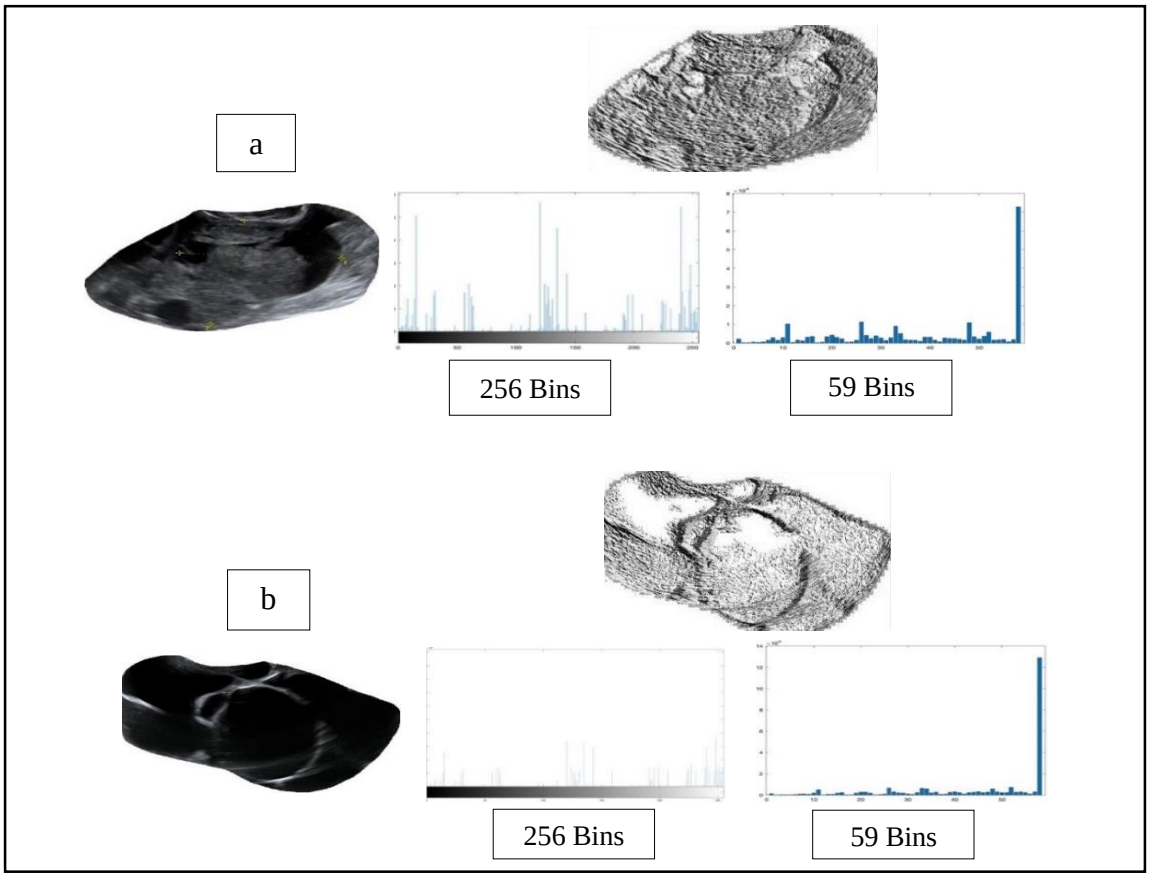


Figure 3- 4: LBP histograms for ultrasound ovarian tumour image of (a) Benign, (b) Malignant.

3.4.3 Geometric based Texture Descriptors

Geometric and shape characteristics of image objects include irregularity of their boundary, their dimensions and orientation, the curviness of textural/structural primitives inside them. Cellular structures created by cancerous cells growing within a suspect tumour region are expected to have a greater deal of irregularity than inside benign tumour. Fractal geometry is a well-established mechanism to understand and categorise geometric shapes that are self-similar at different scales, (Mandelbrot 1983), and hence provides the adequate means of measuring irregularity of cellular network structures. The main objective of fractal geometry is to distinguish dimensionality of 1-dimensional curves when embedded in and fill 2-dimensional shapes. The Fractal dimension (FD) of

a curved image shape is a real number that categorises the geometric complexity of its texture and is different from both intrinsic and Euclidean dimension. It is an indicator of roughness/jaggedness of the tissue texture so that larger FD indicates a more jagged shape texture. In Acharya et al (Acharya, Saba, et al. 2012b) used FD, GLCM and Histogram moments for characterisation and classification of ovarian tumour from 3D ultrasound scans.

The FFGF Frequency domain ultrasound ovarian image texture descriptor, introduced by (Khazendar, Shan 2016), is an interesting low dimensional feature vector that can be characterised as a geometric descriptor. Texture features are extracted from the centralised Fourier transform spectrum of the image through a process of binarization using thresholding high-frequency coefficients (as indicators of significant geometric texture primitives spread throughout the spatial domain image). At a certain threshold, this process results in an elliptical shape bright cluster, the dimensions of which form the FFGF descriptor. The FFGF and its modified versions were used for automatic classification of B-mode Ultrasound ovarian scans into Benign and Malignant cysts.

Interestingly, groups of ULBP codes that have the same number of 1's are obvious examples of geometric texture descriptors that have been used with success in analysing natural images for the detection of image tampering, (Asaad 2017).

This section revealed a variety of image texture feature models and descriptors, of different dimensionalities, some have been already used in analysing natural as well as biomedical images including ultrasound ovarian scan images. In the next chapter, we shall describe several of these descriptors and test their performances in predicting ovarian cysts abnormalities.

3.5 Existing Computer-based Ultrasound Ovarian Tumour Diagnosis

Gynaecologists have long recognised the benefits of analysing US images in detecting and diagnosing ovarian abnormalities, and developed schemes/rules to manually identifying morphologic features observed in B-mode images like unilocular cysts, multilocular cysts, solid tumours, fluid, papillary projections, internal wall structure and acoustic shadows. Moreover, Doppler ultrasound images were adopted to gain vital information about blood flow. Together, these indicators are useful to determine the severity of the cyst. The International Ovarian Tumour Analysis (IOTA) developed several scoring systems, through a refinement process, on using sonographic features. These models are the logistic regression models (LR1, LR2) (Timmerman, Van Calster,

et al. 2010), the risk of malignancy (RMI) (Tingulstad, et al. 1996), Simple rule (Timmerman, Ameye, et al. 2010), and most recently, the ADNEX model (Van Calster, et al., 2015; Kaijser, 2015). While the use of a scoring system helps to improve the test performance, according to (Gramellini, et al. 2008) the existence of several scoring systems can lead to inconsistencies in clinical decision emerging as a result of a number of factors most prominently variation in examiners expertise. Experienced ultrasound examiners use additional demographic information when estimating the types of tumours, while less experienced examiners may struggle to acquire the necessary ultrasound morphology information. Unfortunately, there is no consensus regarding the correct categorisation of morphology information. Moreover, a high percentage of tumours do not conform strictly to some or all the stated rules, and not all masses yield relevant information. Finally, the scoring systems work well with tumours that are easily classifiable using pattern recognition, but less well with those that are not (Timmerman, Ameye, et al. 2010). These challenges, as well as the high cost to healthcare systems, are among the motivations for research into the use of computerised tools to assist in the process of decision making.

Advances in scientific research into cancer have led to the emergence of the hypothesis that links the onset of cancer ovarian tissue with changes in the texture features in ultrasound scans of the ovary. This, together with the rapid increase in computational power, explains the recent interest in investigating and developing texture feature descriptors and designing algorithms to testing their viability for discriminating benign from malignant tumours. Moreover, the rapid emergence of Machine learning as a source of efficient tools for analysing and discovering characterising patterns in massive amounts of natural and medical images.

Having described the categorisation of image texture features and illustrated the plethora of such ultrasound image relevant texture, we shall know review existing research approaches that exploit the great opportunities provided of AI and Machine learning to develop automatic diagnostic support tools. The various schemes, reviewed in this section, have a common structure in that all follow more/less the pipeline described above, while the main differences relate the choice of texture descriptor, or combination of them, the classifier employed in determining the performances of these descriptors, the experimental dataset used, and the main objectives of the proposed scheme including type of tissue targeted, the scanning modality and the abnormality of interest. Most existing schemes use manual segmentation of the region of interest. For the pre-processing step,

generally no explicit discussion is made, but the problem of speckle-noise is dealt with by one of the above de-noising schemes, applied to all trained and tested images regardless of the intensity of the speckle noise.

Research has been conducted to evaluate the power of different US image texture features in discriminating between Benign and Malignant tumours, classifying different types of ovarian tumours as well as detecting pregnancy abnormalities such as miscarriages. 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, and the statistical texture features were extracted using 64 features-based histogram moments, along with 56 features extracted from GLCM in four directions. Average classification accuracy of 86.90% was achieved in identifying different types of benign tumours.

In (Acharya, Saba, et al. 2012b), investigated the performance of a Decision Tree (DT) classifier to classify ovarian tumours from 3D ultrasound ovarian scan images. They extracted Higher-Order Spectra (HOS) features from the image that combines the Fractal Dimension with GLCM and Histogram moments. Due to the fact that this mixed texture feature is of high dimension, the authors apply dimension reduction based on feature selection. The images used were only from 20 patients (10 benign and 10 malignant), and the training benefitted from an abundance of images (1000 images from each benign and malignant tumours). The proposed scheme performed well and achieved 95.1% accuracy. In an alternative investigation, the authors used advanced imaging and a probabilistic neural network (PNN) classifier to achieve an even higher accuracy of 99.8%. This method involved using Hu's invariant moments and Gabor transform parameters and entropies for feature extraction. However, again it was based on the same 20 patients were used to derive the 2,600 images in this study (Acharya, Mookiah, et al. 2014), and 3D colour Doppler images were used instead of B-mode ultrasound.

The well-known SVM classifier was investigated, in (Khazendar, et al., 2014; Khazendar, et al., 2015) with Local Binary Pattern (LBP). and adapted to determine the strength of LBP histogram texture descriptor in distinguishing Benign from Malignant ovarian masses from a B-mode ultrasound scan of the ovary. The authors, not only determined the performance of the LBP but also introduced a simple but very useful concept of confidence in the prediction made during fresh testings. This concept differs from the apriori statistical confidence interval, in that it is defined at every single decision

using the relative distance from the support vector hyperplane. Three discrete confidence levels (low, medium, and high) are defined so that the further away a feature is from the SVM hyperplane the higher the confidence is. When their LBP scheme was trained and tested with a dataset of 177 patients and 187 ultrasound images, classification accuracy rates of 69%, 81%, and 90% were obtained at three different confidence levels (low, medium, and high), respectively. However, the average accuracy of 77%, at 95%, high confidence level, though welcome, was achieved at the expense of a large number of potential tumours misclassified.

In another investigation using the same dataset in (Khazendar, Shan 2016), the authors extracted the geometric-based texture feature FFGF from the Fast Fourier Transform (FFT) domain. The study has different steps, first, pre-process the ultrasound images used a combination of noise removal methods, the second step was to transform an image into FFT frequency domain and compute its spectrum, the third step, binarise the FFT power spectrum, using a manually selected threshold. The final step determines the best ellipse fit of the highlighted shape in the centre of the binary spectrum image from which 3 parameters were used as the feature vector representation of the tumour. The FFGF 3-dimensional record of the ellipse are: (***major axis***, ***major axis***, ***area***). The SVM classifier was used to classify between different types of ovarian tumours with an average accuracy of 85%. The idea will be later modified in this thesis to extract other ellipse parameters; more details will be given in chapter 6.

Another ovarian tumour diagnostic machine learning-based scheme was developed, in (Aramendia-Vidaurreta, et al, 2016) extracts and combine several texture-based features to be subjected to certain feature selection process followed by dimension reduction. This scheme incorporates the patient's age and uses a Neural Network Learning (NNL) approach for classification. When trained and tested on 145 patients (106 benign and 39 malignant tumours), this scheme achieved an accuracy rate of 98.78 %. The combination of feature selection, from several ones, dimension reduction, and use of NNL classifier make it difficult to determine the effect of each feature on decision-making. Moreover, no consideration is given to the confidence level of decisions. To some extent, this scheme works as a black-box decision-maker, which also characterises the recent trends in using Deep Learning Convolutional Neural Networks (CNN) medical diagnostic schemes.

A recent survey conducted by (Kourou 2015), which shows the importance of ML and how it's significant to classify the cancer patients into high or low-risk groups has led many research teams, from the biomedical and the bioinformatics field, to study the

application of machine learning (ML) methods. Therefore, these techniques have been utilized as an aim to model the progression and treatment of cancerous conditions. In addition, the ability of ML tools to detect key features from complex datasets reveals their importance. A variety of these techniques, including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs) and Decision Trees (DTs) have been widely applied in cancer research for the development of predictive models, resulting in effective and accurate decision making. Even though it is evident that the use of ML methods can improve our understanding of cancer progression, an appropriate level of validation is needed in order for these methods to be considered in the everyday clinical practice. The authors in (Kourou 2015) present a review of recent ML approaches employed in the modelling of different types of cancer such as (breast, colon, lung, prostate and *etc*.). The predictive models discussed in (Kourou 2015) are based on various supervised ML techniques as well as on different input features and data samples. Given the growing trend on the application of ML methods in cancer research.

3.7 Summary

This chapter aimed to review the literature on existing work on the computational aspect of analysing ovarian ultrasound scans to develop machine learning algorithms that support gynaecologist in their decisions with regards to diagnosing benign and malignant ovarian cysts. Having outlined the pipeline framework for such machine learning tools, and guided by a basic understanding of how cancerous cells grow and create its own cellular network within human body tissue it became clear why most research publications in the field focus on developing classification schemes rely on the analysis of textural features extracted from ultrasound scans of the ovary. We reviewed existing work on the initial steps prior to texture features and explained that we would rely on manual segmentation by the medical experts at this stage and we shall deal with de-noising in an adaptive way that depends on the effects of speckle noise on the extracted texture feature. We then reviewed image texture features relevant to the intended task which revealed a rich source of such schemes on natural image analysis that could and have been incorporating some existing CAD systems that have been developed for distinguishing between benign and malignant masses. This review has led to modifying the categorisation of texture descriptors into three non-exclusive categories: Statistical-based, transform-based and geometric-based descriptors. To some extents, this categorisation and the review of existing CAD schemes provide the road map for our work in the next 3 chapters.

Chapter 4

Texture-based Ultrasound image Analysis for Ovarian Tumor classification

This chapter is the first of 3 chapters that are aimed to develop automatic texture-based informative machine learning tool(s) suitable for use in support of correct preoperative diagnostic decision making of ovarian masses. The reviews and discussions covered in the previous chapter helped to draft a roadmap for conducting our research investigations to guide a realistic and effective implementation of a standard pipeline of procedures. Although each component of the pipeline is essential for the success of developing the intended tool, the most crucial component is that of choosing well-performing image features to be extracted from the ROI. For example, the challenge of automating segmentation of cysts regions can be deferred at this stage. On the other hand, the appropriate pre-processing step should be designed adaptively according to the effect of image quality and speckle noise on the performance of the extracted features. Moreover, the choice of classifiers is an aftermath component. Accordingly, this chapter will be focusing on developing a set of feature vectors that singularly or together through fusion provide effective separation of the images of benign cysts from those of malignant ones.

4.1 Introduction

According to the evolving understanding of the way carcinogenesis results in changing cellular networks the texture of cysts, the starting assumption of our research is that these changes will be manifested in the Ultrasound scan images of the ovaries. To a large extent, this hypothesis is backed by the trends in existing research in this multi-disciplinary field. Hence, our research will be based on the analysis of a variety of texture descriptors extracted from B-mode ultrasound ovarian scan images. Pattern recognition research in the non-medical domain has led to a variety of models for quantifying texture features in natural images, and it is natural to use these models to extract texture features from medical images to derive a pool of texture descriptors. However, instead of focusing on the best performing texture features in any area of pattern recognition, we shall first select many known and well-understood list of texture features and try to estimate their performances.

Ultrasound imaging is an imaging modality and application that is extensively used in medicine and related areas of research. It offers various benefits in comparison with other modalities, such as a non-invasive manner of usage, mobility, low cost, and the potential for real-time imaging. However, the process of image transmission and acquisition causes images to distorting due to the noise. Noise and artefacts are generated in different types of medical imaging modalities, each causing specific and prominent signal and image deterioration. Ultrasound produces coherent imaging that is degraded by a category of noise termed as speckle noise, which can seriously affect image quality. Such output will subsequently alter the medical interpretation and performance of various computer-assisted techniques. Moreover, the efficacy of the image properties may also hinder the different phases of feature extraction, analysis, segmentation, recognition and more, thus rendering them unconvincing. However, at this stage, we do not know what would be the effect of speckle noise on each of our yet to be selected features or if the effect is uniform on all features. Therefore, we shall not apply any pre-processing procedure but will come back to this in the next chapter.

4.2 Ultrasound Image Texture Feature Descriptors

In chapter 3, we described and categorised commonly used image texture features, but here we shall describe the mathematical formulae for some known features that facilitate the computation of their descriptors. The chosen texture features may be classified by more than one of the models/categories reviewed in chapter 2.

4.2.1. Statistics Histogram Properties

Assume the image is a function $f(x, y)$ with x and y space variables, where y, $x=0,1,\ldots,N-1$ and $y=0,1,\ldots,M-1$. The function $f(x, y)$ can take discrete values $i = 0,1,\ldots,G-1$. In this case, G refers to the total number of intensity levels within the ultrasound image. The histogram will show the number of pixels in the entire image at each intensity level, as follows:

$$H(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(i,j) \tag{4.1}$$

Where $\delta\ (i,j)$ is the Kronecker delta function

$$\delta(i,j) = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{4.2}$$

This type of histogram that displays the pixel intensity levels provides a quick summary of the statistical data from the original image. Grey-level histograms are calculated based on single pixels, providing first-order image information. It is possible to calculate an estimate of the probability density of occurrence of the intensity levels by diving the $h(i)$ by the total number of pixels contained within the ultrasound image.

It is relatively easy to compute the histogram from the original image. The overall shape of the histogram can provide some details of the image itself. For example, if the image has low contrast, the histogram will appear narrowly distributed. Furthermore, in an image that has a specific region with low contrast, surrounded by regions with a wide range of intensities, the histogram will appear bimodal. Data from the histogram can help to determine the first-order statistical features of the ultrasound image. In some cases, texture can be characterised from the central moments (Papoulis and Pillai 2002) that are derived from the histogram. If i is a random variable representing an image pixel intensity, $p(i)$ is the histogram of the intensity levels in an image, and G is the total number of pixel in the image then the statistical properties of a histogram can be computed using the mean, variance, entropy, kurtosis, skewness, and energy (Materka, Strzelecki and others 1998) and together, they define the statistical properties of an image.

- **Mean**: - is measured by averaging all the individual pixel intensities. The mean will determine the level of brightness in the image.

$$\mu = \sum_{i=0}^{G-1} ip\,(i) \tag{4.3}$$

- **Variance** - is the variability in the levels of pixel intensity from the mean.

$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 P(i) \tag{4.4}$$

- **Skewness**: - The histogram skewness describes the distribution of the histogram; especially if one of the sides has a longer tail. A positive or negative skewness value indicates that the intensity levels of the pixels are heavily weighted above or below the mean, whereas a skewness value of 0 indicated that the histogram is evenly distributed.

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 P(i) \tag{4.5}$$

- **Kurtosis**: - measure the height of the histogram peak and the angle of incline towards the peak.

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 P(i) - 3 \tag{4.6}$$

- **Entropy**: - describe the randomness of the pixel intensities.

$$H = -\sum_{i=0}^{G-1} p(i) \log_2[p(i)] \tag{4.7}$$

- **Energy**: - The energy is also referred to as the uniformity of energy and angular second moment.

$$E = \sum_{i=0}^{G-1} [p(i)]^2 \tag{4.8}$$

4.2.2 Local Binary Pattern LBP

Local Binary Patterns (LBP) is an image transformation that maps each pixel intensity value into a byte that conveys the order relation between the pixel and its immediate neighbours in a specific orientation. In its simplest form, the LBP byte of an image pixel in position (x,y) is calculated bit-by-bit from left to right, using the order relation between the boundary pixels of a *3x3* window centred at (*x,y*) starting with the top left corner pixel and moving in a clockwise direction: if the border pixel value is greater than or equal to that of (*x,y*) then the corresponding bit is set to 1 else it is set to 0 (Ojala, Pietikäinen and Harwood 1996). Figure 4-1 below shows a hypothetical *3x3* image block to illustrate this concept and displays an image with its LBP transforms.

Figure 4- 1: (a) Single LBP code for 3x3 block (P=8, R=1), (b) Image with its LBP transform.

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 defining a similarity relationship among the patterns. Grouping the different patterns result 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:

I. Simple LBP (256) bins:

An LBP that has a high level of similarity is an indication of the equality of the patterns. In this case, the LBP mapping provides a histogram that shows the frequency of each 8-bit binary code. Each 8-bit code refers to a decimal grey-level value. The histogram is formed using 256 bins of the LBP map from the image.

II-Uniform LBP (59 bins):

The identification of similar aspects in local binary patterns is determined according to the number of transitions between 0 and 1 in the patterns as the primary mechanism of uniform LBP. Among the diverse patterns produced by the dominant approach in this regard is the so-called "uniform" pattern $LBP_{u2P,R}$ (P represents the pixels proximal to radius R); a local binary pattern must include no more than two bitwise transitions from 0 to 1 or from 1 to 0 (with the equivalent bit string being deemed circular). Examples of uniform patterns are 00000000 (0 transitions) and 01111110 (2 transitions), while 11101101 (4 transitions) and 01011011 (6 transitions) are non-uniform patterns. Every uniform pattern is associated with a different output label in uniform LBP (see Figure 4-2), whereas just one label is allocated to all the patterns that are not uniform. Hence, patterns can be mapped based on the number of distinct output labels (Pietikainen, et al. 2011).

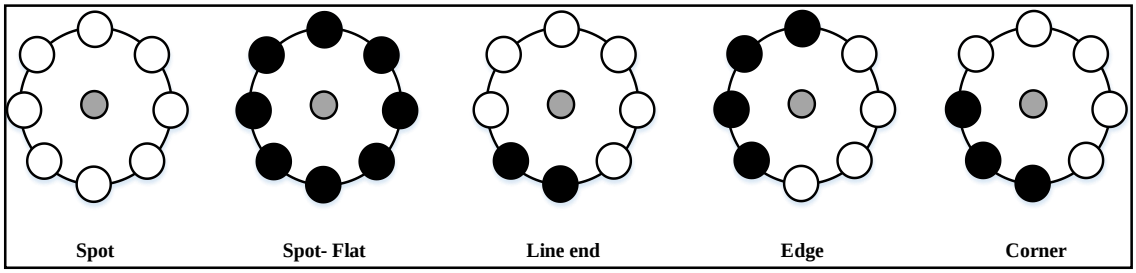


Figure 4- 2: Different texture primitives detected by LBP.

Removal of non-uniform patterns can be justified for several reasons. First of all, uniformity is exhibited by the majority of local binary patterns in natural images. Secondly, evidence suggests that the stability of uniform patterns is higher, in other words, they have a lower susceptibility to noise (Pietikainen, et al. 2011). Thirdly, the number of potential LBP labels is considerably reduced when only uniform patterns are taken into account, and fewer samples are needed to reliably estimate their arrangement.

Its main applications include textural classification, image extraction and face image analysis (Lei, Zhao, & Guo, 2015; Pietikainen, Hadid, Zhao, & Ahonen, 2011).

4.2.3 Gabor Filter

Gabor filters are designed for detecting specific frequency information around an image pixel/region in a given direction, and there strong and credible arguments by computer vision experts that the frequency and orientation of Gabor filters as similar to the human vision system (Daugman 1985). A two-dimensional Gabor filter is based on the Gabor Wavelet Transform. It acts as a Gaussian kernel in two variables modulated by a sinusoidal plane wave on the x-axis and thus consists of imaginary and real orthogonal components. It allows one to analyse an image, in terms of spatial variation, to build a texture discrimination model (Weldon, Higgins and Dunn 1996), (Ilonen, et al. 2008). In this model, images are transformed by a set of parametrised Gabor filters, applied at a different set of scales and orientations. The imaginary and real orthogonal components of Gabor filter (*gf*) are defined for the pair of scale and orientation (l, k) by:

$$gf_{(l,k)}(m,n) = \frac{f^2}{\pi\gamma\eta}\exp(-x'^2\frac{f^2}{\gamma^2} + y'^2\frac{f^2}{\eta^2})\exp(j2\pi x') \qquad (4.9)$$

Where f is the sinusoidal factor frequency, θ is the direction of the normal to the parallel stripes of the Gabor function, x' = m cos (θ) + n sin (θ) and y' = −m sin (θ) + n cos (θ), γ is the sharpness along the major axis X, η is the sharpness along the minor axis Y, (m, n) represents the dimensions of the image, *l* is the scale, k signifies the orientation factor, and λ = η/γ represents the aspect ratio of the Gaussian. Normally a set of Gabor filters are used. These are Gaussian filters of various sizes that are modulated using sinusoidal plane waves set to multiple orientations. Kernel size is relative to the dimension of the Gaussian filter. That said, kernel size must be, at a minimum, six times the standard deviation with the nearest odd number chosen. The Gobor wavelet representation *xl, k* of an image is the twisted distortion of the image with a family of Gabor wavelets. (Ilonen and Kälviäinen 2005) (Ilonen, et al. 2008).

4.2.4 Histograms of Oriented Gradient

Features found within images may be identified using the histogram of oriented gradients (HOG). It is a method developed to quantify the gradient orientation in specific sections of an image. What guides HOG is the arrangement of local intensity gradients, which is used in item feature in an image even if there is no accurate data about equivalent gradient or edge positions. Successfully using this approach involves creating many cells within

the image window, each containing an aggregate of the local one-dimensional histogram of gradient directions for age orientations over the cell pixels. The resulting representation is the summation of histogram entries. Furthermore, it is best to ensure the representations are contrast normalized before employing local responses to improve invariance illumination, shadowing size (blocks) and use of results to achieve block cell. The normalized descriptor blocks are the same thing as the HOG descriptors. The human detection chain is obtained by using a dense and overlapping HOG descriptive grid for tiling the detection window and using the combined feature vector. (Dalal and Triggs 2005). As shown in Figure 4-3, the HOG your extraction algorithm operates through a series of steps. In the first step, the input image is segmented into the small interlinked cell, and for each cell a histogram of gradient directions for the pixels is generated

- **Step1**: The gradient orientation is used to delineate every cell in angular bins;
- **Step 2**: The weighted gradient is contributed to by the pixel of every cell to its equivalent angular bin;
- **Step 3**: Neighbouring cells cluster into spatial areas known as blocks, which constitute the foundation for histogram aggregation and normalisation;
- **Step 4**: The result of the block histogram is the group histogram that has been normalised, and the descriptor is indicated by a block histogram set.

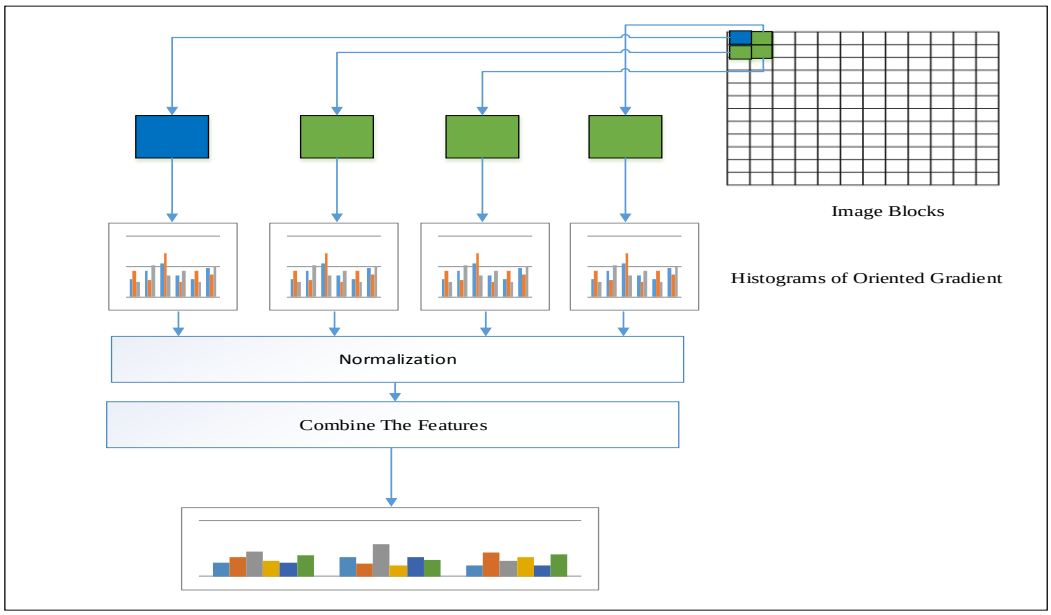


Figure 4- 3: HOG process.

4.2.5 Fractal Dimension (FD)

In the 1960's, Benoit Mandelbrt introduced the concept of fractals to model geometric

shapes that are self-similar at different scales (Mandelbrot 1983) . The large volume of subsequent research has demonstrated that such seemingly chaotic shapes can be generated through a deterministic iterative mathematical formula. The well-known example of the attractive shapes of Julia sets and their widespread use in the art are testaments to the success of fractal models of shapes. The interest in fractal geometry originated in how to distinguish dimensionality of 1-dimensional curves when embedded in and fill 2-dimensional shapes, and this led to the coining the Fractal dimension (FD) to distinguishes from both ***intrinsic*** and ***Euclidean*** dimension. The FD of an image shape is a real number that categorises the geometric complexity of its texture. It is an indicator of roughness/jaggedness of the shape texture so that larger FD indicates a more jagged shape texture. Different methods have been developed to calculate the FD of objects for use in segmentation and image texture analysis, but the ***Box Counting*** scheme originally suggested by Mandelbrt is the most common and simple practical one adopted for computer vision. It works, by (1) binarizing the image, (2) cover and count the RoI with boxes of different sizes, and (3) plot the sequence of points (box size, No. of boxes) in 2-dimension and the slope of the best fit line is the FD of the image object. A linear regression model, which uses a logarithmic scale, is used to fit the line and determine FD. Figure 4-4 shows a fractal curve covered different box sizes at distinct levels, varying from lowermost to uppermost. Here, r_i is the ratio of structure size to the total, and $N(r_i)$ is the number of units covering the structure. Accordingly, FD is defined, (Mandelbrot 1983) , by the limit:

$$FD = {\log(N(r))}\Big/{\log(1/r)} \tag{4.10}$$

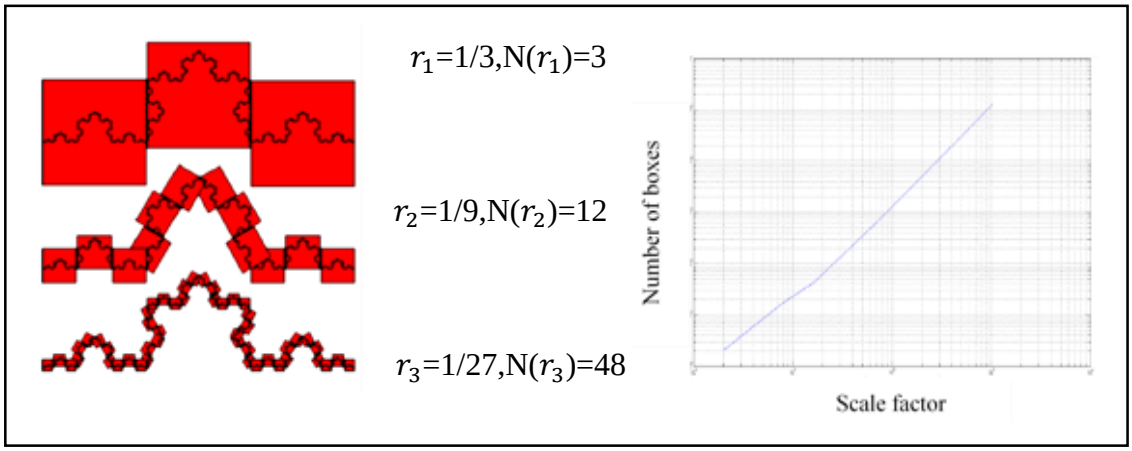


Figure 4- 4: FD example (scale factor (1/r) and number of boxes (N(r))).

4.2.6 Hu's Invariant Moments (IM)

In 1962, (Hu 1962) Ming-Kuei Hu introduced a set of 2-dimensional moment invariants of images as the visual information theory model for image analysis. He showed that image recognition schemes based on these invariants are truly position, size and orientation independent, and are sufficiently flexible to learn almost any set of patterns. These invariants are defined in terms of doubly parametrised functions defined on the data domain in terms of weighted averages of pixel strengths. In total, seven invariant moments are deemed to be sufficient, including invariance under rotation, translation, and scaling. For any digitally sampled M × M grey image $f(x,y)$, define a set of two-dimensional moments

$$mpq = \sum_{x=0}^{x=M-1} \sum_{y=0}^{y=M-1} (x)^p . \quad (y)^q \ f(x,y) \ \ p,q = 0,1,2,3,4 \ldots\ldots \qquad (4.11)$$

Some of these values are similar to the moment values extracted from image histograms. Hu's has shown that the double moment sequence $\{m_{pq}\}$ is uniquely determined by $f(x,y)$; and conversely, $f(x,y)$ is uniquely determined by $\{m_{pq}\}$. The *centralised* moments of $f(x,y)$ are defined by translated by an amount (a, b) as follows:

$$\mu_{pq} \quad = \sum_{x}\sum_{y} \ (x+a)^p \ .(y+b)^q f(x,y) \qquad (4.12)$$

The central moments m'_{pq} of $f(x,y)$ *d*efined as the value of μ_{pq} with a = - $\overline{x}$ and b = - $\bar{y}$, i.e.

$$\mu_{pq=} \sum_{x}\sum_{y} (x-\bar{x})^p \ . (y-\bar{y})^q f\ (x,y) \qquad (4.13)$$

Where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$,

It is important to note that central moments are independent of their orientation or position. However, there is a need to normalise these moments, prior to use for texture analysis. For $p,q = 0,1,2 \ldots$, the normalised central moments are defined by the expressions $\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$, where $\gamma = \frac{p+q}{2} + 1$.

Finally, Hu's moments-based texture feature vector consists of seven parameters defined in terms of the normalised centralised moments of the image $f(x,y)$, as follows:

$\boldsymbol{m1} = (\eta_{2,0} + \eta_{0,2})$.

$$\boldsymbol{m2} = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2.$$

$$\boldsymbol{m3} = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2.$$

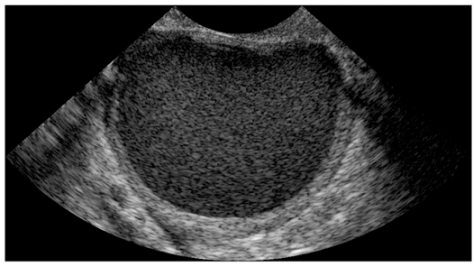
$$\boldsymbol{m4} = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{0,3} + \eta_{2,1})^2.$$

$$\boldsymbol{m5} = (\eta_{3,0} + 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2}) + [(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2] + (3\eta_{2,1} + \eta_{0,3})(\eta_{0,3} + \eta_{2,1})\left[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})\right].$$

$$\boldsymbol{m6} = (\eta_{2,0} - \eta_{0,2})\left[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{1,2})^2\right] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{0,3} + \eta_{2,1}).$$

$$\boldsymbol{m7} = (3\eta_{2,1} - 3\eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2] - (\eta_{3,0} - 3\eta_{1,2})(\eta_{0,3} + \eta_{2,1})[(3\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2] \quad (4.14)$$

In Figure 4.5, an example can be seen. The original image, its 45 grades rotated image version and its half sized image version, together with the seven Hu moments.



	ORIGINAL	ROTATED 45º	HALF-SIZED
hu1	-2,43E-01	-2,43E-01	-2,43E-01
hu2	-1,30E+00	-1,30E+00	-1,30E+00
hu3	-2,48E+00	-2,48E+00	-2,48E+00
hu4	-3,11E+00	-3,11E+00	-3,11E+00
hu5	-6,49E+00	-6,49E+00	-6,49E+00
hu6	3,78E+00	3,78E+00	3,78E+00
hu7	5,96E+00	5,96E+00	5,96E+00

Figure 4- 5: Hu's moments.

4.3. Performance Evaluation of Texture-based Schemes

To evaluate the performance of a designed diagnostic model, besides selecting a classifier we need to have a sufficiently large experimental dataset of benign and malignant Static 2D US ovarian scan images to train the chosen classifier. The dataset must include sufficient and near equal number of both benign and malignant tumours. We also need to use a sensible conventional training/testing protocol that determines the ratio of training to testing samples as well as the criteria for selecting these sets. This section is devoted to describing these issues and to the various choices.

4.3.1. The Experimental Dataset

Ultrasound images that were used in this study were obtained from the IOTA research (Timmerman, Ameye, et al. 2010). A total of 232 women patients gave consent to be included in the study and had ovarian tumours that were surgically removed between November 2005 and November 2013. All ovarian tumours were given a histological diagnosis. A total of 242 (2D B-mode) scan images (138 benign and 104 malignant tumours) were generated (Astraia software gmbh, Germany) at the Department of Gynaecological Ultrasonography, Campus Gasthuisberg, KU Leuven, Belgium. In May 2012, the University of Buckingham's School of Science & Medicine Ethics Committee granted ethical approval. Table 1 shows the histology classes of all the 242 images. (see Table 4.1).Image analysis was restricted to the gray scale portion of the image. For image processing, the MATLAB (Matlab R2017 MathWorks, Natick, Massachusetts, USA) software was used. The number of benign tumours (138) is noticeably larger than the number of malignant ones (104). To resolve this imbalance, we select randomly m (<150) images from each class.

Table 4- 1: Histopathology of ovarian masses.

Number of images	**Benign Tumours (138)**
27	Mature Teratoma
24	Mucinous Cystadenoma
16	Endometrioma/endometriosis
5	Functional Cyst
12	Ovarian Fibroma
26	Serous Cystadenoma
16	Serous Cystadnofibroma
12	(1 abscess, 2 Brenner tumour, 2 Multilocular peritoneal inclusion cyst MPIC, 2 Mucinous Cystadnofibroma, 1 subserous adenomyoma, 2 hydrosalpinx, 2 simple cyst)
Malignant Tumours (104)	
18	Borderline mucinous tumour
11	Borderline serous tumour
38	Serous cyst/adenocarcinoma
6	Mucinous cyst/adenocarcinoma

12	Endometrioid adenocarcinoma
19	(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, 2 Immature Teratoma, 1 endometrial cancer 1 AIDS related lymfoma)

4.3.2. Evaluation Protocols

We use two protocols for training and testing: Leave-one-out (L1O) and the 50-50 balanced training and testing sets. For each protocol, we first use the Randsample function in MATLAB to select a total of m=150 images out of the 242 images with 75 Benign and 75 Malignant images. To reduce the chance of biasness we repeat each experiment 30 times.

1. For each L1O protocol experiment, each pair of Benign-Malignant images are removed we use the remaining 148 images to train the SVM classifier and use the trained model to classify the pair of isolated images and record the success & failure cases. Due to a large number of possible pairs (5625 in this case), it is customary to repeat this for a sufficient number of pairs (75 in our case).

2. For each 50-50 protocol experiment, we randomly select 37 Benign and 37 Malignant images for training the SVM classifier and use the remaining 76 mix of images to test the performance of the trained model counting the number of success and failure.

4.3.3. Performance Measures

For each protocol, the accumulated success and failure cases in each experiment will be used to obtain the performance of the individual feature vector schemes for that experiment. This is done as follows:

Let TP (true positive) = No. of correctly diagnosed malignant masses,

FP (false positive) = No. of benign masses incorrectly classified as malignant,

TN (true negative) = No. of benign masses correctly classified, and

FN (false negative) = No. of malignant masses incorrectly classified as benign.

Then, performance is represented by the following 3 rates:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (4.15)$$

$$Sensitivity = TP/(TP + FN) \quad (4.16)$$

$$Specificity = TN/(TN + FP) \quad (4.17)$$

Here, Sensitivity is the frequency of appropriately classifying diseased individuals, and Specificity is the frequency of appropriately classifying healthy individuals.

4.4. Experimental Results and Discussion

In this section, we present the results of the experiments conducted to evaluate the performance of each texture features, in discriminating between Benign and Malignant cyst, by training and testing two classifiers (SVM and kNN) with the images in the experimental dataset for the two different protocols. Each experiment will be repeated 30 times to reduce dependence on the training/testing choices, and the average accuracy of these experiments provide the estimated performance of the chosen texture feature(s).

4.4.1. Performance of Single Texture-Feature based Scheme

Having repeated each of the described experiments 30 times, the performance of the 7 proposed texture-based schemes are presented in terms of the average and standard deviation of accuracy, sensitivity, and specificity rates of the 30 experiments. Table 4-2 below, displays the performances of the 7 texture-based feature diagnostic schemes for both protocols when the binary SVM classifier is deployed to discriminate between benign and malignant ovarian masses.

Table 4- 2: Classification results based on SVM classifier.

Performance Rate / Feature vector	Accuracy mean, (Stdev)		Sensitivity mean, (Stdev)		Specificity mean, (Stdev)	
	L1O	50-50	L1O	50-50	L1O	50-50
LBP (256 bins)	74.46% (3.076)	79.41% (3.570)	75.23% (2.839)	80.57% (2.358)	73.88% (3.957)	78.25% (1.258)
7-Moments	74.94% (2.736)	76.60% (2.025)	75.34% (1.168)	78.60% (1.685)	74.54% (2.723)	74.60% (2.638)
Statistics Histogram	79.88% (4.097)	80.79% (2.076)	80.98% (1.914)	81.64% (2.574)	78.78% (1.605)	79.94% (2.754)
HOG	80.11% (2.074)	82.64% (0.997)	80.25% (3.714)	82.96% (1.068)	79.98% (2.254)	82.33% (1.145)
ULBP (59 bins)	81.29% (1.016)	83.12% (1.002)	82.26% (1.627)	83.89% (1.574)	80.33% (2.745)	82.35% (1.247)
Fractal Dimension	81.57% (1.963)	83% (1.120)	82.18% (2.675)	83.25% (1.987)	80.97% (2.599)	82.76% (1.458)
Gabor Filter	84.60% (0.132)	86.70% (0.020)	85.05% (0.322)	86.85% (0.003)	84.68% (0.742)	86.55% (0.017)

Overall, the results show in Table 4-2 using SVM classifier that each of the texture features has reasonably high discriminating characteristics and demonstrates the viability of using artificial intelligence and machine learning as an effective supporting tool of

digital health. The lowest accuracy being well above 70% and the fact that all features were extracted directly from the ROI, without enhancing/de-noising the images, clearly support this statement about texture feature analysis.

These results reveal a variation in the performance of the different schemes with average accuracy in the ranges of (74.46 - 84.60) for the L1O protocol and (75.36- 86.70) for the 50-50 protocol. Moreover, for each scheme, there is no significant difference between specificity and sensitivity rates attained with either protocol. This is more obvious in the case of the Gabor scheme. In ascending order of accuracy, these schemes are in the order (LBP256, 7-Moments, Stat-Hist, HOG, ULBP59, FD, and Gabor filter) with respect to the L1O protocol. This order changes slightly, with little significance, when the 50-50 protocol is deployed. On the other hand, for all features the use of the 50-50 protocol results in better performance than the use of the L1O protocol but the improvement is only statistically significant in the case of the Gabor filter feature as result of the fact that the standard deviation values (for all the 3 rates) are remarkably low relative to the other features.

The relatively low performance of the LBP256 and the 7-Moments schemes can be attributed partially to the global nature of these statistical histogram-based features. However, the variation in mage size may have a significant impact on the adopted distance/similarity function between histograms and statistical parameters. The LBP256 is sensitive to small changes, while moments are inherently location-dependent, more sensitive to several geometric transformations (rotation, translation, scaling) and non-geometric transformations (smoothing). Although LBP features statistical capture distribution of local variation in intensity, the scheme cannot reflect global spatial information. The histogram of the LBP256 suffers from the presence of significant redundancies reducing its discrimination in different classes, and this may explain the significant improvement in the performance of the LBP59 scheme. Similarly, the improved performance of the Statistics histogram scheme over the 7-Moments may be due to the presence of dependency between the different moment values.

The next improved performance is achieved by the HOG texture-feature scheme albeit marginally in the case of L1O protocol and modestly with the 50-50 protocol. This is probably due to the fact that the HOG feature expresses the location of intensity gradients that help identify edge directions in different image blocks. It provides additional useful information on the shape and appearance of the structures within the image. Further marginal performance improvement is achieved by two almost equally performing

schemes: the ULBP59 and the FD schemes. Interestingly, both features are linked to the geometry of image shapes/objects. While FD represents the degree of texture irregularity in terms of self-similarity of objects at different scales, the ULBP59 codes are linked to distinctive texture geometries (flat local regions, edges and corners). Finally, the Gabor filter scheme outperforms all other schemes in terms of accuracy and balanced level of errors in the two classes with respect to both protocols. More importantly, all the 3 performance rates do not seem to depend on the selection of the training set of samples. This is demonstrated by the extremely small standard deviations in all cases and protocols this due to that the same amount of images are classified correctly for benign and malignant cases. Perhaps, these results reinforce the widely accepted assertion that Gabor filters provide a credible mathematical model of the Human Vision system.

Next, we repeated the same experiments, but this time, we replaced the SVM with the kNN (where k value =1) classifier. Table 4.3 shows the overall classification accuracy rates achieved by each single feature scheme using kNN classifier. We have tested the same feature methods using kNN classifier based on both experimental protocols, as shown in Table 4-3.

Table 4- 3: Classification results based on kNN (k value =1) classifier.

Performance Rate / Feature Vector	Accuracy mean, (Stdev)		Sensitivity mean, (Stdev)		Specificity mean, (Stdev)	
	L1O	50-50	L1O	50-50	L1O	50-50
LBP (256 bins)	71.79% (3.637)	73.33% (3.814)	72.67% (2.145)	75.62% (4.653)	71.28% (3.010)	73.04% (3.236)
7-Moments	69.06% (2.733)	72.03% (3.876)	70.11% (2.231)	72.21% (2.609)	68.01% (2.278)	71.85% (3.324)
Statistics Histogram	70.82% (3.157)	74.20% (1.453)	70.89% (2.989)	74.49% (2.121)	70.76% (3.001)	73.91% (2.985)
HOG	72.60% (3.034)	75.20% (1.450)	72.79% (1.354)	76.31% (1.023)	71.93% (1.869)	74.10% (2.235)
ULBP (59 bins)	74.67% (2.952)	77.08% (1.007)	75.90% (2.364)	77.87% (1.465)	73.44% (2.250)	76.29% (1.812)
Fractal Dimension	75.17% (3.080)	78.39% (1.008)	76.29% (3.814)	79.01% (1.001)	73.79% (3.528)	77.77% (1.654)
Gabor Filter	79.41% (1.693)	82.33% (0.897)	80.21% (1.042)	82.78% (1.078)	78.62% (1.367)	81.89% (1.147)

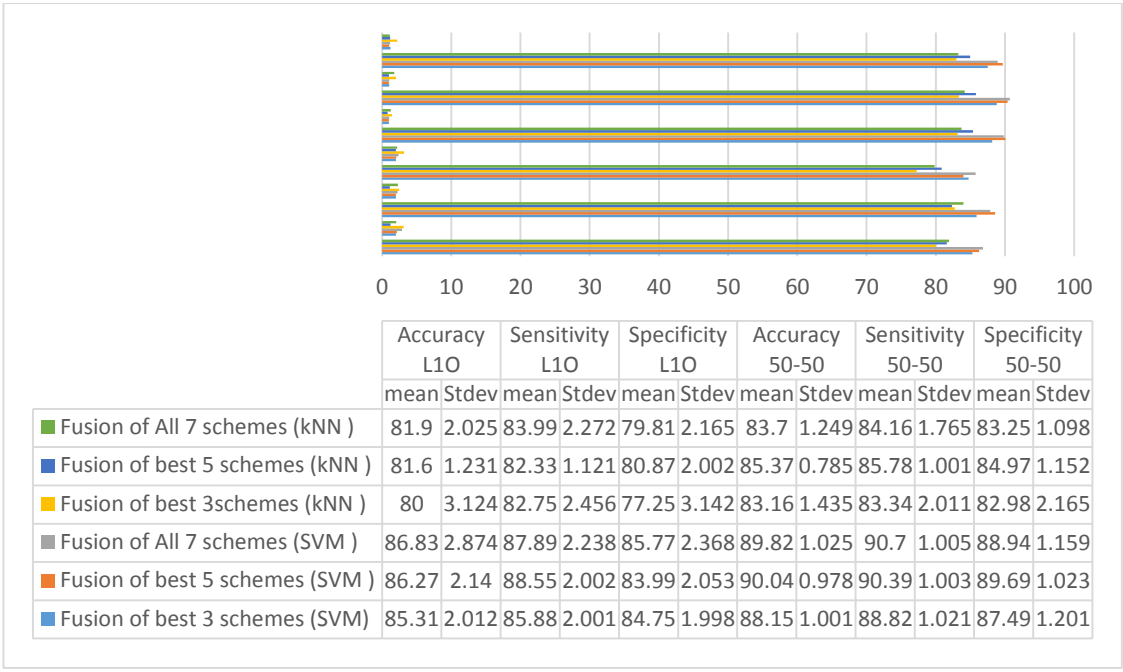
A close examination of the results displayed in Table 4-2 and Table 4-3, shows that the SVM consistently outperforms the kNN classifier for both experimental protocols on all types of features for classifying benign and malignant tumours. Considering that most forms of the feature vector are of relatively high dimension, the impressive performance of SVM in high dimensional feature spaces is an example of its superiority over the kNN

classifier. Furthermore, SVM outperforms kNN for the statistics histogram properties and seven moment's features that are both of low dimensionality. US scan images of malignant tumours tend to contain more detailed and complex structures compared to images of benign tumours. These results also indicate that fusion may provide even more improved accuracy. Therefore, the SVM and kNN classifiers are used based on various seven feature extraction schemes.

4.4.2 Decision-based Fusion of Multiple Schemes

All the features investigated above are based on different mathematical models of ovarian tumour structure and/or texture. The performances of their diagnostic schemes varied and though mostly satisfactory, it is not clear if they make similar decisions or not. This raises the possibility of combining these different schemes (perhaps using different classifiers) to compensate for individual weaknesses in a collaborative manner to improve decision credibility akin to seeking multiple medical expert opinions when diagnosing life-threatening diseases.

The well-known and widely practised approach of combining different features/classification schemes is that of ***fusion*** which can be done at different levels (e.g. features level, score level, and decision level). Feature level fusion is not realistic in our case, because of significant differences in the structure of the features and normalising them would be highly complex. Fusion at the score level is more appropriate but again requires some kind of normalisation of the different scores, which would be difficult in our case due to the differences in the structure, and complexity of the feature vectors. Fusion at the decision level is by far the easiest to implement and at this stage could provide better informative advice on the suitability of the fusion approach. Here, we will assess the performance of decision level fusing of the above, different feature methods. Although there is no requirement on using different classifiers for a different feature vector, we will confine our investigation to using the same SVM and kNN (where k value =1) classifiers for all features. Moreover, instead of developing a complicated fusion rule, we will opt for the fusion by simple majority rule. This means that the final decision is based on the highest vote for the two classes by the various schemes, and therefore an odd number of feature schemes are to be fused. The use of the simple majority rule can be justified by the fact that the various schemes had comparably reasonable performances. Figure 4-6 presents the results of majority rule fusion of SVM and kNN classification by different odd numbers of features.



	Accuracy L1O		Sensitivity L1O		Specificity L1O		Accuracy 50-50		Sensitivity 50-50		Specificity 50-50	
	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev
Fusion of All 7 schemes (kNN)	81.9	2.025	83.99	2.272	79.81	2.165	83.7	1.249	84.16	1.765	83.25	1.098
Fusion of best 5 schemes (kNN)	81.6	1.231	82.33	1.121	80.87	2.002	85.37	0.785	85.78	1.001	84.97	1.152
Fusion of best 3schemes (kNN)	80	3.124	82.75	2.456	77.25	3.142	83.16	1.435	83.34	2.011	82.98	2.165
Fusion of All 7 schemes (SVM)	86.83	2.874	87.89	2.238	85.77	2.368	89.82	1.025	90.7	1.005	88.94	1.159
Fusion of best 5 schemes (SVM)	86.27	2.14	88.55	2.002	83.99	2.053	90.04	0.978	90.39	1.003	89.69	1.023
Fusion of best 3 schemes (SVM)	85.31	2.012	85.88	2.001	84.75	1.998	88.15	1.001	88.82	1.021	87.49	1.201

Figure 4- 6 : Performance measures obtained using the majority voting fusion using SVM & kNN (k value =1) classifiers.

The experiment confirms the success of the decision-based fusion approach in improving accuracy for both protocols and the 3 fusion configurations over the best performing single scheme. The absolute rate of best improvement for L1O is 2.23 when fusing all 7 schemes while for the 50-50 protocol is 3.44 when fusing best 5 schemes. Moreover, in all cases, the standard deviation values for the 50-50 protocol remain close to 1% for the different rates, while the corresponding values for the L1O protocol are in the range 2-3%. The fusion results confirm that in the case of the 50-50 protocol, the performance of the classifier is less dependent on the choice of the training/testing sets.

However, it is apparent that SVM and kNN tend to make correct predictions for both benign and malignant tumours. However, the SVM outperforms kNN even when features are fused based on a simple method like majority voting. The difference in performance is because the kNN can have poor run-time performance when the training set is huge. It is very sensitive to irrelevant or redundant features since all features contribute to the similarity and classification. Also, kNN is sensitive to noise since it is based on the Euclidean distance.

By examining the ultrasound images and the sensitivity and specificity of each classifier, it is apparent that the SVM and kNN usually predict both benign and malignant tumours accurately. However, the SVM is better than kNN, even when the features are fused using a simple method like majority voting. Also, this experiment, confirm the success of the decision fusion approach as the accuracy improved around 3% over the best

performing scheme. These experiments illustrate a reasonable level of complementarities of these different features in representing tumour tissue texture.

4.5 Summary

We investigated several image texture-relevant feature vectors, extracted from B-mode Ultrasound scan images of ovarian tissue masses, to determine their discriminating powers in relation to distinguishing Benign from Malignant cysts. The results from these investigations are to feed into our ultimate objective of developing automatic informative machine learning tool(s) to be used in support of correct preoperative diagnostic decision making of ovarian masses. In total, seven such texture analysis schemes were developed and used to train SVM classifiers, the performance of 7 such feature was evaluated for a dataset of IOTA scan images of 238 patients, and each and every scheme attained accuracy well above random decision. The Gabor filters, known to its effectiveness in modelling the human vision system, attained the highest stable accuracy of 86.76%. Decision level fusion of different combinations of these schemes resulted in nearly 90% accuracy. These results demonstrate the viability and effectiveness of using AI for the stated task.

In this chapter, we confined our experiments to determine the contribution of each texture descriptor singularly and combined by decision level fusion without mitigating the effects of factors that are known to adversely influence pattern recognition and classification in computer vision. The next chapter will focus on investigating the influence of such factors and means of mitigating their influences on the performance of the above schemes.

Chapter 5

Mitigating Factors Influencing the Performance of the Texture-based Ovarian Tmours Diagnostic Schemes

In the last chapter, we designed, developed and tested the performance of several ovarian tumour schemes using different texture-based feature vectors extracted automatically from the cropped mass regions. Although, in Chapter 3, we highlighted the potentially adverse effect of Speckle noise on the performance of automatic ovarian tumour detection and diagnostic. However, no action was included within the diagnostic schemes of Chapter 4 with regards to noise removal. This decision was based on a lack of clear knowledge of the effect of noise on the performance of each of diagnostic texture-based schemes. In the first part of this chapter, we aim to investigate this link. The marked variation in the size of the ROIs is another factor that may have an adverse influence on the performance of those schemes that involve the use of histogram descriptors.

The chapter is organised so that first in section 5.1, we shall demonstrate that blanket speckle de-noising of all images at the training and testing stages may not enhance the chance of accurate diagnostic decisions for all images. In section 5.2, we define an image parameter that can be computed for any input US Ovarian image, as an indicator to whether the level of speckle-noise in an input image merits or not de-noising prior to feature extraction and recognition. In section 5.3, we shall use this measure to develop an adaptive speckle de-noising schemes and test the performance of the schemes, tested in Chapter 4, when using a blanket de-noising in comparison to using the adaptive de-noising scheme. The second part of this chapter aims to develop a procedure to mitigate the second factor on variation in ROI sizes in experimental datasets and test the effect of the developed procedure on recognition accuracy.

5.1 Effect of Speckle-Noise on the Classification of US Ovary Scans

Medical ultrasound imageries are sensitive to this specific type of noise that is known to reduce image contrast and make boundaries of the different regions in scanned tissues blurred/undetectable. Accordingly, it may have an adverse effect on automatic segmentation, texture feature extraction and possibly diagnostic related classification. While it is essential to incorporate the speckle-noise reduction scheme as a pre-processing

step for our intended texture-based US ovarian image classification, it is essential to quantify the effect of speckle noise on US ovarian scan images.

5.1.1 Speckle Noise in Ultrasound Images

Noise is visible in the image through unwanted variation observed in brightness or colour. One example is electronic noise, which can be caused by the sensors of a digital camera or a scanners' circuitry. Noise can appear because of a slow shutter speed or from light generated through high exposures (Hedrick and Hykes 1989). There are other types of noise, such as salt & paper, quantisation noise, and impulse noise. These types of noise are distinguished by the nature of their distributions, and various filters have been developed for de-noising images. Identifying the nature of the noise is vital when selecting the type of filter to use when rectifying an image.

Medical ultrasound images are known to be subject to a specific type of locally correlated multiplicative noise, known as speckle noise, due to the fact that biological tissues react to incident ultrasound waveform by a combination of scattering and reflection. If a particular reading is below a sound's wavelength, it is identified as scattering. Otherwise, it is a reflection. This variation, like the acoustic waves, is caused by differences in the medium's thickness or compressibility. Once the back-scattered acoustic pulses arrive back to the transducer, they receive acoustic energy with constructive and destructive involvement, which causes the appearance of a granular pattern referred to speckle noise (Dangeti, 2003; Burckhardt, 1978). Speckle noise reduces image quality because it hides and distorts essential features, especially around the edges, and makes image segmentation as well as post-processing operations less reliable and may reduce the quality of image diagnostic, (Zhu, Ni, Li, & Gu, 2009; Xie, Jiang, Tsui, & Heng, 2006).

Level of image degradation in response to speckle noise in different image regions depends very much on the brightness of the region so that higher variation in pixel intensity will be noted in brighter regions than in darker regions. This could be observed simultaneously within an ovarian ultrasound image due to the natural variation of cell types in the scanned tissue as well as the level of tissue solidity in and around tumour regions. Figure 5-1 illustrates the level of speckle-noise effect in different regions of an ultrasound ovarian tumour image.

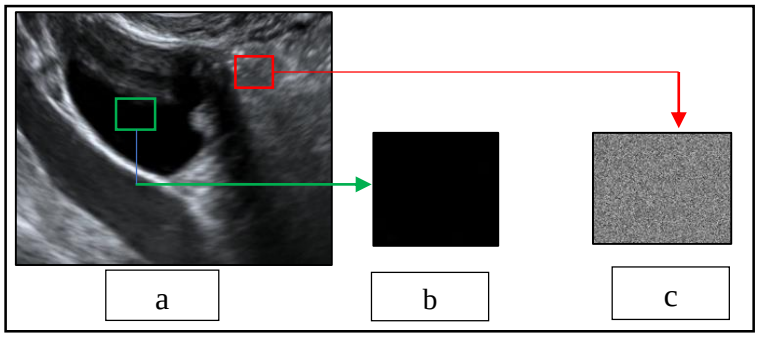


Figure 5- 1: (a) A US image of an ovarian tumour corrupted by different levels of speckle-noise: Low level in part (b), High level in part (c).

5.1.2 Impact of Speckle Noise in Ultrasound Images on Diagnosis

Reducing speckle noise is vital for effective automatic processing and analysis of ultrasound images. It is widely believed that de-noising and improving the quality of ultrasound ovarian tumour images can lead to enhancement of diagnosis. Since we have several texture-based diagnostic schemes and the different textures may have different spatial distribution, then it is essential to determine the effect of speckle-noise removal on each scheme. Hence, we repeated the same set of experiments conducted in Chapter 4, but this time, we applied a well-known speckle-noise reduction on the RoIs of all the training and testing images. We shall now describe this speckle-noise suppression scheme.

Speckle noise is multiplicative and non-Gaussian and hence is generally more difficult to remove than additive noise. A model of multiplicative noise is given by the formula: -

$$y_{ij} = x_{ij+}\, n_{ij} \tag{5.1}$$

Where the observed noisy image (y_{ij}) is the product of the original unknown image (x_{ij}), and (n_{ij}) is the non-Gaussian noise image. Speckle noise is typically presumed to be data-independent, as well as fixed with a unitary mean and unknown noise variance of σ^2 in most applications that incorporate multiplicative noise. A simple logarithmic transformation converts equation (5.1) to an additive noise model:

$$log(\, y_{ij}) = log\big(\, x_{ij}\big) + log(\, n_{ij}) \tag{5.2}$$

The Wiener filter (Wiener 1949) is a linear spatial domain filter but can be applied to (5.2) in two ways: (1) in the Fourier transform (frequency) domain, and (2) in the spatial domain using the mean square technique. The first option is normally useful for both de-noising and de-blurring, whereas the second option is just applied for de-noising. The frequency-domain option requires advanced knowledge of the spectra of noise power and

the original image. However, the spatial domain option requires no prior knowledge as it is founded on a statistical least square principle that minimises the mean squared error between iteratively estimated (x_{ij})'s in equation 5.2.

The statistical properties (e.g. mean, variance, and higher-order moments) within ultrasound images fluctuate significantly from one region to another. Therefore, both global statistics over the entire image and local statistics over the kernel can be important to assess the success of de-noising. The global and local statistics post-Wiener filtering is characterised by:

$$Y_{ij} = \overline{K} + \frac{\sigma_k^2}{\sigma_k^2+\sigma^2}\,(K_{uv} - \overline{K}) \qquad (5.3)$$

Where *Yij* indicates the de-speckled image, the local mean is $\overline{K}$, σ_k^2 is the local variance over the kernel *K*, *Kuv* refers to the pixel values in *K*, and σ^2 is the global variance. Therefore, the filter output is equal to the local mean if the central pixel is equivalent to the local mean. Otherwise, it is modified by both terms. The Wiener filtered image is finally subjected to the exponential transform to invert the logarithm operation and produce a de-noised image. In Table 5-1, we present the performance of the 7 texture-based schemes with and without this blind de-noising procedure for both protocols.

Table 5- 1: Performances of 7 texture feature methods with blind pre-processing and without pre-processing based on SVM & kNN (k value =1) classifiers.

SVM Classifier						
Performance Rate / **Feature Vector**	**Accuracy mean, (Stdev)**		**Sensitivity mean, (Stdev)**		**Specificity mean, (Stdev)**	
	L1O	**50-50**	**L1O**	**50-50**	**L1O**	**50-50**
7-Moments No Pre-processing	**74.94% (2.736)**	**76.60% (2.025)**	**75.34% (1.168)**	**78.60% (1.685)**	**74.54% (2.723)**	**74.60% (2.638)**
7-Moments Blind de-noising	**75.62% (3.118)**	**76.90% (2.878)**	**77.04% (3.201)**	**76.88% (2.785)**	**74.20% (3.210)**	**76.92% (2.965)**
Statistics Histogram No Pre-processing	**79.88% (4.097)**	**80.79% (2.076)**	**80.98% (1.914)**	**81.64% (2.574)**	**78.78% (1.605)**	**79.94% (2.754)**
Statistics Histogram Blind de-noising	**79.67% (4.147)**	**79.80% (3.899)**	**81.36% (3.085)**	**80.72% (3.394)**	**77.98% (4.101)**	**78.88% (3.985)**
HOG No Pre-processing	**80.11% (2.074)**	**82.64% (0.997)**	**80.25% (3.714)**	**82.96% (1.068)**	**79.98% (2.254)**	**82.33% (1.145)**
HOG Blind de-noising	**79.44% (2.214)**	**80.36% (2.751)**	**80.87% (2.852)**	**81% (2.798)**	**78.02% (2.927)**	**79.73% (2.698)**
FD No Pre-processing	**81.57% (1.963)**	**83% (1.120)**	**82.18% (2.675)**	**83.25% (1.987)**	**80.97% (2.599)**	**82.76% (1.458)**
FD Blind de-noising	**82.02% (3.758)**	**83.45% (3.241)**	**84.25% (3.475)**	**85.90% (3.028)**	**79.79% (4.025)**	**81.01% (3.423)**
Gabor Filter	**84.60%**	**86.70%**	**85.05%**	**86.85%**	**84.68%**	**86.55%**

No Pre-processing	**(0.132)**	**(0.020)**	**(0.322)**	**(0.003)**	**(0.742)**	**(0.017)**
Gabor Filter	**84.03%**	**86.44%**	**84.98%**	**86.93%**	**83.08%**	**85.95%**
Blind de-noising	**(2.042)**	**(2.734)**	**(2.121)**	**(2.378)**	**(2.352)**	**(2.565)**
LBP (256 bins)	**74.46%**	**79.41%**	**75.23%**	**80.57%**	**73.88%**	**78.25%**
No Pre-processing	**(3.076)**	**(3.570)**	**(2.839)**	**(2.358)**	**(3.957)**	**(1.258)**
LBP (256 bins)	**75.51%**	**80.03%**	**77.13%**	**80.22%**	**73.01%**	**79.85%**
Blind de-noising	**(2.344)**	**(2.864)**	**(2.876)**	**(1.234)**	**(2.001)**	**(2.212)**
LBP (59 bins)	**81.29%**	**83.12%**	**82.26%**	**83.89%**	**80.33%**	**82.35%**
No Pre-processing	**(1.016)**	**(1.002)**	**(1.627)**	**(1.574)**	**(2.745)**	**(1.247)**
LBP (59 bins)	**82.11%**	**83.33%**	**83.04%**	**83.98%**	**81.19%**	**82.68%**
Blind de-noising	**(2.009)**	**(2.211)**	**(2.112)**	**(2.342)**	**(2.234)**	**(2.009)**
kNN Classifier						
7-Moments	**69.06%**	**72.03%**	**70.11%**	**72.21%**	**68.01%**	**71.85%**
No Pre-processing	**(2.733)**	**(3.876)**	**(2.231)**	**(2.609)**	**(2.278)**	**(3.324)**
7-Moments	**72.71%**	**74.40%**	**73.22%**	**73.85%**	**72.21%**	**74.96%**
Blind de-noising	**(2.865)**	**(2.878)**	**(2.006)**	**(2.785)**	**(3.474)**	**(2.965)**
Statistics Histogram	**70.82%**	**74.20%**	**70.89%**	**74.49%**	**70.76%**	**73.91%**
No Pre-processing	**(3.157)**	**(1.453)**	**(2.989)**	**(2.121)**	**(3.001)**	**(2.985)**
Statistics Histogram	**73.89%**	**75.59%**	**74.76%**	**76.82%**	**73.02%**	**74.36%**
Blind de-noising	**(4.147)**	**(3.750)**	**(3.085)**	**(3.084)**	**(4.101)**	**(3.702)**
HOG	**72.60%**	**75.20%**	**72.79%**	**76.31%**	**71.93%**	**74.10%**
No Pre-processing	**(3.034)**	**(1.450)**	**(1.354)**	**(1.023)**	**(1.869)**	**(2.235)**
HOG	**76.46%**	**77.37%**	**77.55%**	**77.91%**	**75.38%**	**76.83%**
Blind de-noising	**(2.074)**	**(2.647)**	**(3.094)**	**(2.948)**	**(3.103)**	**(3.051)**
FD	**75.17%**	**78.39%**	**76.29%**	**79.01%**	**73.79%**	**77.77%**
No Pre-processing	**(3.080)**	**(1.008)**	**(3.814)**	**(1.001)**	**(3.528)**	**(1.654)**
FD	**78.81%**	**80.41%**	**81.72%**	**82.25%**	**75.90%**	**78.57%**
Blind de-noising	**(3.272)**	**(3.657)**	**(3.884)**	**(3.855)**	**(3.995)**	**(3.980)**
Gabor Filter	**79.41%**	**82.33%**	**80.21%**	**82.78%**	**78.62%**	**81.89%**
No Pre-processing	**(1.693)**	**(0.897)**	**(1.042)**	**(1.078)**	**(1.367)**	**(1.147)**
Gabor Filter	**80.84%**	**81.55%**	**81.58%**	**83.25%**	**80.11%**	**81.85%**
Blind de-noising	**(2.005)**	**(2.045)**	**(2.101)**	**(2.104)**	**(2.302)**	**(2.276)**
LBP (256 bins)	**71.79%**	**73.33%**	**72.67%**	**75.62%**	**71.28%**	**73.04%**
No Pre-processing	**(3.637)**	**(3.814)**	**(2.145)**	**(4.653)**	**(3.010)**	**(3.236)**
LBP (256 bins)	**74.21%**	**76.21%**	**74.38%**	**76.44%**	**74.04%**	**75.98%**
Blind de-noising	**(2.003)**	**(1.221)**	**(2.312)**	**(1.675)**	**(2.344)**	**(2.004)**
LBP (59 bins)	**74.67%**	**77.08%**	**75.90%**	**77.87%**	**73.44%**	**76.29%**
No Pre-processing	**(2.952)**	**(1.007)**	**(2.364)**	**(1.465)**	**(2.250)**	**(1.812)**
LBP (59 bins)	**78.21%**	**79.26%**	**78.98%**	**79.98%**	**77.44%**	**78.55%**
Blind de-noising	**(2.987)**	**(1.541)**	**(2.300)**	**(1.112)**	**(2.123)**	**(1.543)**

The results presented in Table 5-1, indicate that the performance of each of feature schemes do not differ clearly with and the without de-noising scenarios. For some features, the "with blind de-noising" scheme marginally outperforms the "without de-noising" scheme, but for other features, the results go the other way around. Even when there is a noticeable difference in the performance, the relatively large standard deviations of accuracy render such difference as not being statistically significant. This observation may be miss-interpreted to naively conclude that there is no genuine need for de-noising.

However, this may indicate that for each feature some subset of the images benefit from de-noising and contribute to better classification, while other images become more difficult to classify after de-noising and thereby cancelling out the earlier improved accuracy. This is an incentive to develop an ***adaptive speckle de-noising*** scheme that could be applied only when needed. This often means the need for a quantitative indicator to be used for determining the need for de-noising or otherwise. The previous observation about the variation in the level of noise caused degradation between different tissue regions, raises the following question: do we need a single adaption indicator to be obtained from the whole RoI or a regional-based indicator? In the next section, we should try to answer this question, to meet an overall objective of improving the accuracy of tumour classification and describe the approach we follow in defining the adaption indicator(s).

5.2 Trainable Systems for Adaptive Speckle Noise Removal

To design adaptive procedures, one needs to identify the image parameter(s) that can be used as indicators of the level of noise present in the image. Due to the fact we are interested in improving the diagnostic accuracy of texture-based schemes, it is natural to examine the values of some texture feature parameter values in a few cases where we ground truth is known. The limited amounts of experiments revealed that the statistical parameters of Kurtosis and Skewness of the intensity distribution in the RoIs provide a clear separation of images that need de-noising from those that do not. Images dominated by dark regions tend to have significantly/reasonably lower Kurtosis/Skewness values than images that are dominated by solid regions (i.e. bright), (e.g. see Figure 5-2).

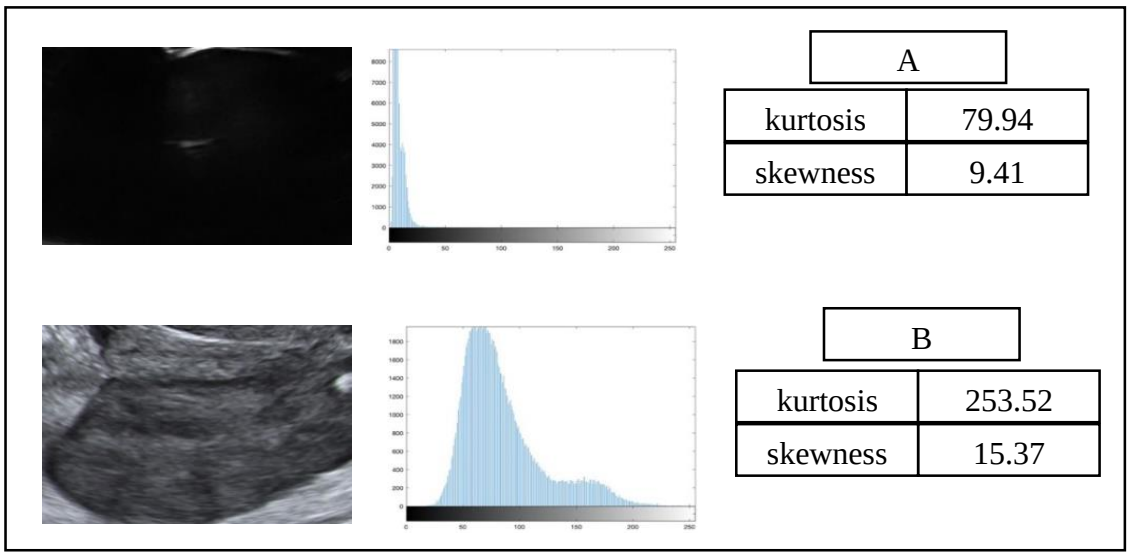


Figure 5- 2: Kourtosis & Skewness of tumour tissue images (A) little solidity; (B) predominantly solid.

Finding good thresholds for the adaptive procedure, however, is complicated by the difficulty in automatically deciding whether the RoI is predominantly dark or predominantly bright. Several attempts on selecting an appropriate threshold revealed the futility of this approach. We, therefore, opted for using a trainable system that uses these two parameters in classifying images into a class of images that need speckle de-noising and a class that does not need it.

Although, this approach seems to be sensible and the corresponding adaptive de-noising scheme may improve the accuracy of some or all the previously developed and tested texture-based diagnostic schemes it does not exploit the knowledge that RoI's may include subblocks of solid as well as non-solid tissues resulting in cancelling out the effect of de-noising. Accordingly, we decided to use two models of a trainable system of adaptive de-noising:

1. **Model 1:** Training based on the two parameters computed over the whole RoI, and adaptively applying/bypassing the de-noising scheme on the entire RoI. The pre-processing in this model is of global nature.
2. **Model 2:** Training based on the two parameters computed in reasonable size subblocks the RoI, and adaptively applying/bypassing the de-noising scheme on each block. Here the pre-processing is applied locally with only a few blocks are de-noised, if any.

Next, we describe these two models in more details and present their effects on the performance of each of the texture-based diagnostic schemes.

5.2.1 Model 1: Adaptive Global Speckle Noise Reduction

This model works in two stages. In stage 1, we train an SVM (see (Cortes and Vapnik 1995)) classifier on a dataset of 2-dimensional vectors (kurtosis, skewness) extracted from the RoI of images from two groups: class1 where the scanned ovarian tissue has no or little solid components; and class 2 where the scanned tissue is mostly brighter, showing complex structure, indicative of presence of solidity. For this, we manually selected 10 images from each class, and these images were not used in testing. Table 5-2 displays two samples, from each class, that were used for training.

Table 5- 2: Samples of a class 1 and class 2 RoIs of ovarian ultrasound images.

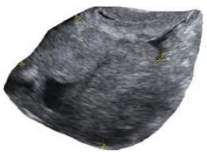
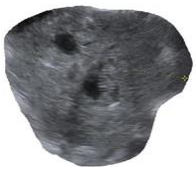
Class1 ROI	Class 2 Solid ROI
	
	

Figure 5-3 displays the flow diagram of the training system of this model. In the testing stage, any new input region of interest is classified as class 1 or class 2, based on its (kurtosis, Skewness) vector representation, and if class 1 label is output then no de-noising is applied, otherwise is will pre-processed for a reduced level of speckle noise.

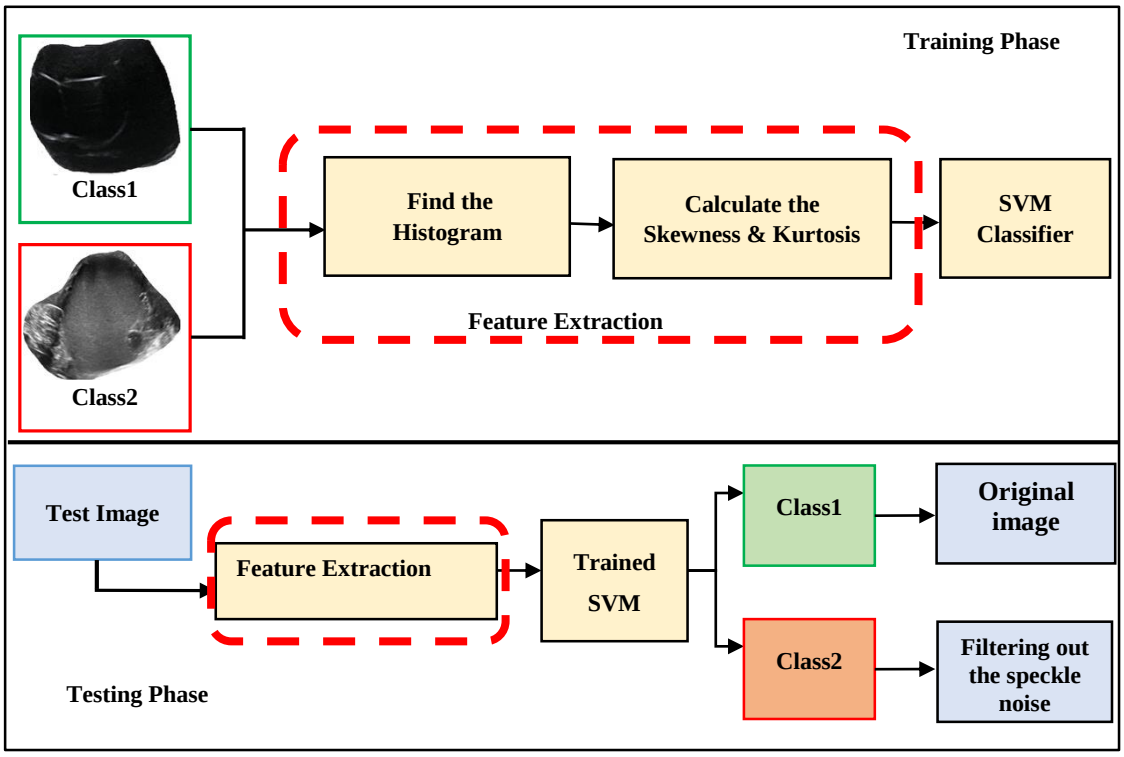


Figure 5- 3: Model1: Speckle noise based on the whole image.

The Wiener filter was implemented only on the image that had a solid ROI. The Wiener filter was applied in different kernel sizes of 3x3, 5x5, 7x7 and 9x9. The filter of 3x3 kernel size, in particular, was found to be effective in retaining the edges and other properties to a certain extent and hence provided the best visual enhancement of the image and texture classification.

5.2.2 Model 2: Localised and Adaptive Speckle Noise Reduction

In some instances, RoI images include both solid and non-solid textured regions (see Figure 5-4. In these cases, a global de-noising may add some unintended artefacts while no de-noising may leave some degradations and blurred edges and, in both cases, the diagnostic decision may become unreliable. We, therefore, propose developed a localised and adaptive de-noising model could be built by dividing the images based into blocks to capture the local information (different level of speckle-noise) and act accordingly.

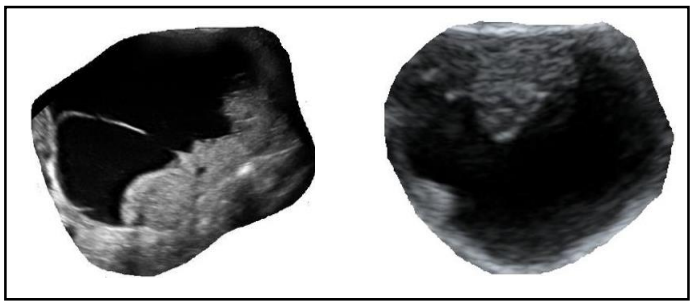


Figure 5- 4: Ovarian tumour samples illustrate the different level of speckle-noise .

Again, we use a training system to build Model 2 adaptive speckle de-noise procedure. In the training phase, 100 block samples from each class with optimal size 30×30, were collected (see Figure 5-6). A histogram was conducted on each, followed by the calculation of static measurements of skewness and kurtosis based on the histogram readings. These feature vectors were used to train an SVM pattern classifier to differentiate between in-class (no solid tissue) and out-of-class (solid tissue) patterns. Figure 5-5 shows samples of blocks used for training.

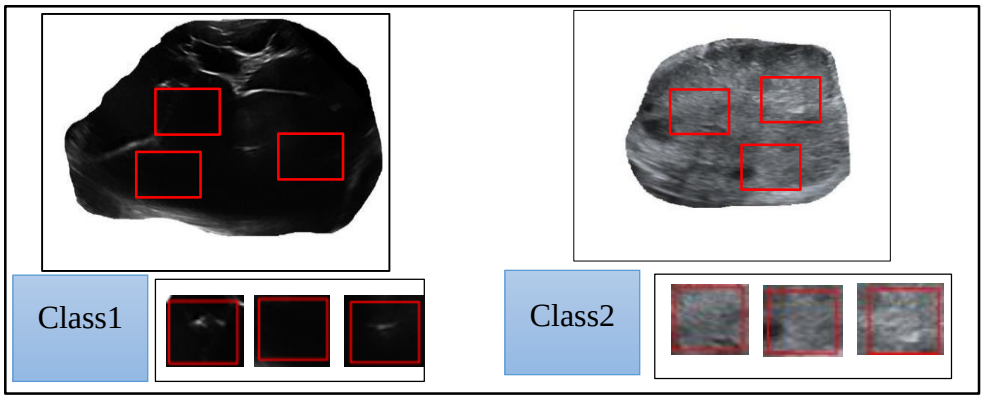


Figure 5- 5: Samples of images used to train Model 2.

In the testing phase, the aim was to capture solid areas in out-of-sample images. The size of the block was fixed over an image, and the trained classifier is used to determine which patterns show the solidity in the block. At each window position, the same set of variables were used in the training step, and they were fed into the classifier. The classifier output

determined whether the block needs to be processed or not. In the next section, we present the results of experiments conducted to test the performance of the texture-based diagnostic schemes post using the adaptive speckle de-noising with both models.

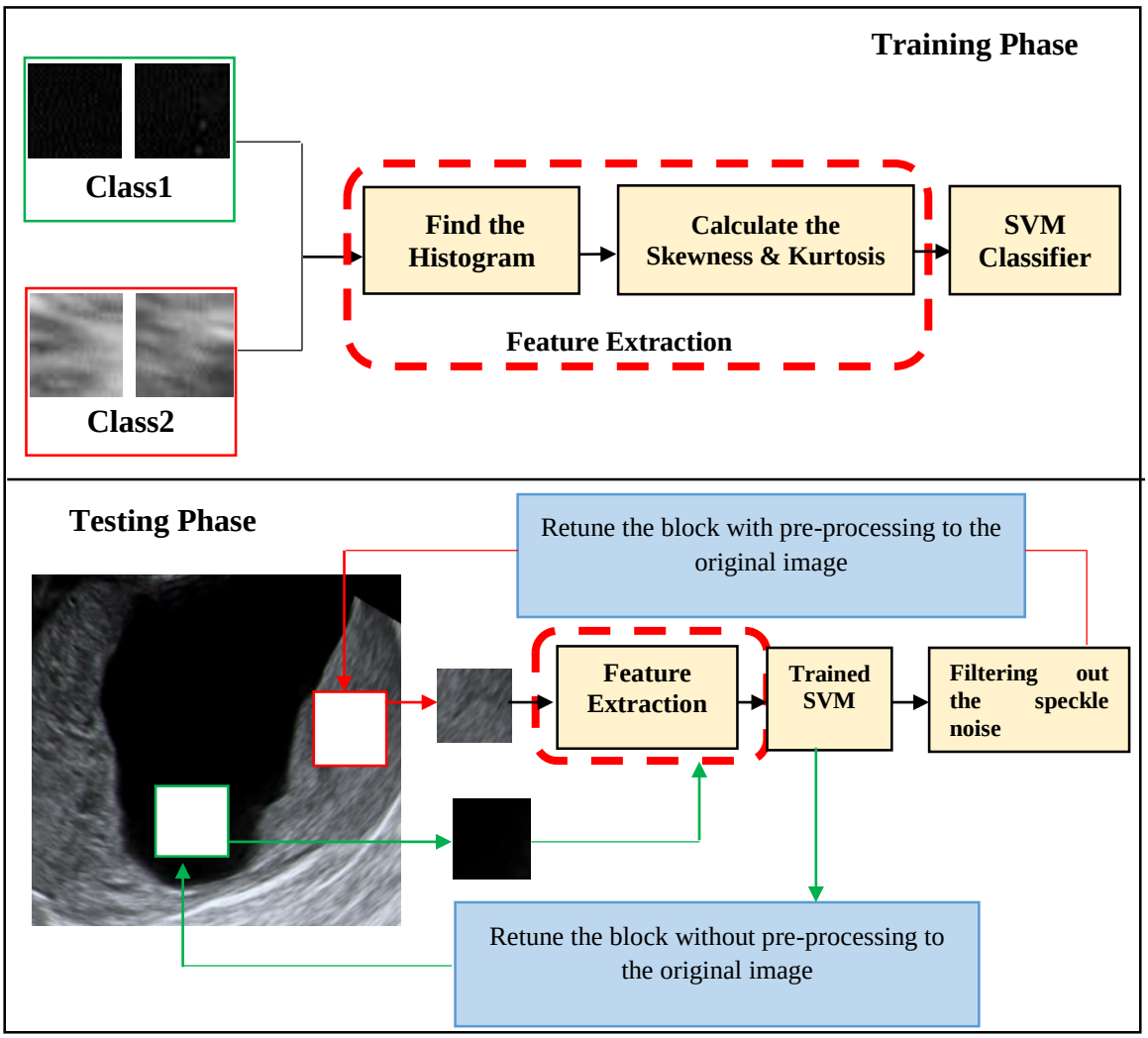
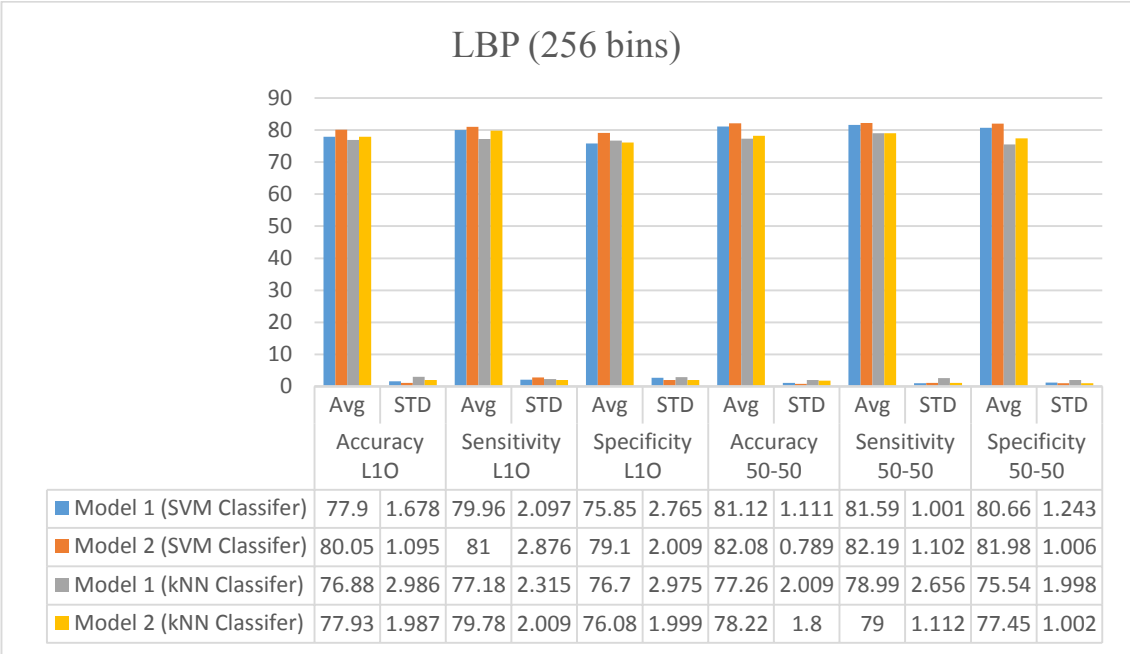


Figure 5- 6: Model2: Adaptive localized Speckle noise suppression.

5.3 Adaptive Pre-processing – Experiment and Discussion

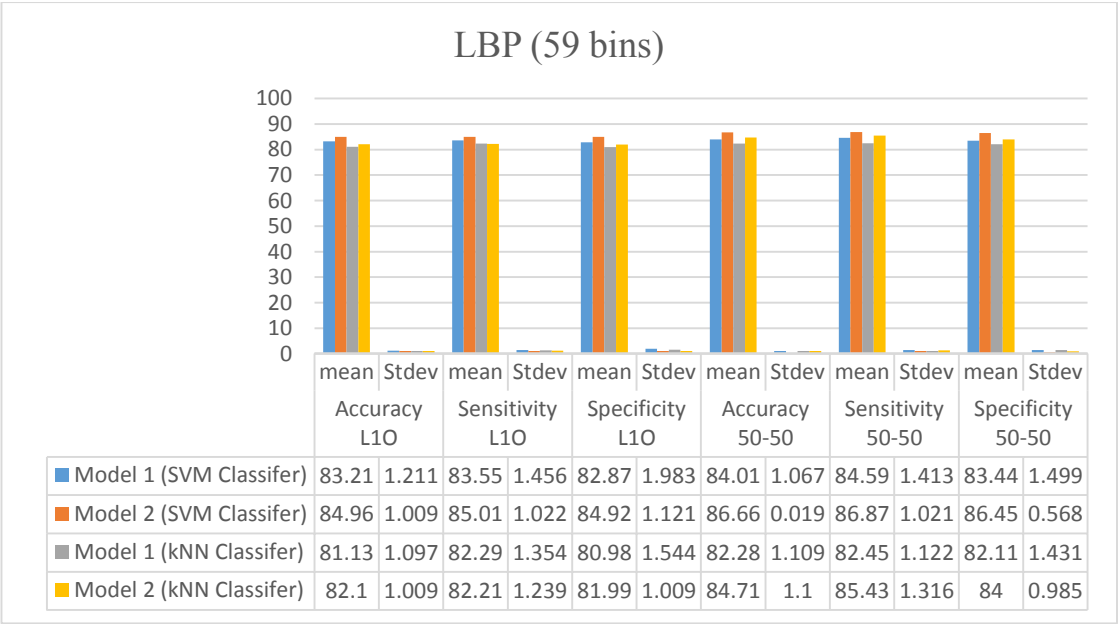
We conducted a large number of experiments to test the effect of the two models of adaptive speckle denoising on pre-processing input Ultrasound Ovarian tumour regions for improved quality and minimal noise distortion suitable for diagnostic decisions using our proposed texture-based schemes. In these experiments, 7 texture-based feature schemes, introduced in chapter 4, were tested to evaluate the effects of the adaptive de-noising models on the diagnostic performance of both SVM and kNN classifiers for both adopted experimental protocols (L1O & 50-50, see chapter 4). The results of these experiments are to be compared with the cases of applying no pre-processing and blind de-noising.

We first compare the performance of two related features (LBP (256 bins) and LBP (59 bins)) post de-noising by model 1 with post de-noising with model 2 for both SVM and kNN classifiers. These results are shown in Figure 5-7 and Figure 5-8, below.



	Avg Accuracy L1O	STD Accuracy L1O	Avg Sensitivity L1O	STD Sensitivity L1O	Avg Specificity L1O	STD Specificity L1O	Avg Accuracy 50-50	STD Accuracy 50-50	Avg Sensitivity 50-50	STD Sensitivity 50-50	Avg Specificity 50-50	STD Specificity 50-50
Model 1 (SVM Classifer)	77.9	1.678	79.96	2.097	75.85	2.765	81.12	1.111	81.59	1.001	80.66	1.243
Model 2 (SVM Classifer)	80.05	1.095	81	2.876	79.1	2.009	82.08	0.789	82.19	1.102	81.98	1.006
Model 1 (kNN Classifer)	76.88	2.986	77.18	2.315	76.7	2.975	77.26	2.009	78.99	2.656	75.54	1.998
Model 2 (kNN Classifer)	77.93	1.987	79.78	2.009	76.08	1.999	78.22	1.8	79	1.112	77.45	1.002

Figure 5- 7: Performance of the LBP (256) based scheme post adaptive de-noising.



	mean Accuracy L1O	Stdev Accuracy L1O	mean Sensitivity L1O	Stdev Sensitivity L1O	mean Specificity L1O	Stdev Specificity L1O	mean Accuracy 50-50	Stdev Accuracy 50-50	mean Sensitivity 50-50	Stdev Sensitivity 50-50	mean Specificity 50-50	Stdev Specificity 50-50
Model 1 (SVM Classifer)	83.21	1.211	83.55	1.456	82.87	1.983	84.01	1.067	84.59	1.413	83.44	1.499
Model 2 (SVM Classifer)	84.96	1.009	85.01	1.022	84.92	1.121	86.66	0.019	86.87	1.021	86.45	0.568
Model 1 (kNN Classifer)	81.13	1.097	82.29	1.354	80.98	1.544	82.28	1.109	82.45	1.122	82.11	1.431
Model 2 (kNN Classifer)	82.1	1.009	82.21	1.239	81.99	1.009	84.71	1.1	85.43	1.316	84	0.985

Figure 5- 8: Performance of adaptive pre-processing for the LBP (59) texture scheme.

These results show that model 2 yield better performance than model 1 for both features and classifiers. The LBP (59 bins) scheme outperform the LBP (256 bins) scheme for both classifiers, and for both schemes, the SVM yield better accuracy than the kNN classifier. The use of adaptive de-noising improves the accuracy of both features and

classifiers in comparison to both no de-noising and blind de-nosing schemes as presented in Table 5-1 above. In summary, one can conclude at least for these two features we must use adaptive de-noising.

Opting for adaptive pre-processing techniques according to blocking allows for vital components in ovarian tumour diagnosis like solid areas (i.e. rich textures) to be identified. However, model 1 is advantageous and suitable for images that only contain solid textures. However, it is not suitable for images that include other types of textures (i.e. solid and not solid). The results led to the second model being proposed. Based on these results, the recommended algorithm according to the blocking model (2) resulted in a notable increase in subjective image quality and did not produce detectable artefacts. The algorithm displayed comparatively better performance than the model (1). When conducted the same experiments for the other 5 schemes, we got a similar pattern of performance to the LBP based schemes in relation to model 1 and model 2. However, we only present in Table 5-3 only when tested the other 5 feature extraction methods using Model 2 only for the other 5 methods, i.e. statistics histogram parameters, Gabor filter, histograms of oriented gradients (HOG), fractal dimension (FD) and Hu's invariant Moments using SVM & kNN (where k value =1) classifiers.

Table 5- 3: Performance of 5 texture-based diagnostic schemes post model 2 adaptive de-noising based on SVM and kNN (k value =1) classifiers.

SVM Classifier (Model 2)						
Performance Rate / Feature vector	**Accuracy mean, (Stdev)**		**Sensitivity mean, (Stdev)**		**Specificity mean, (Stdev)**	
	L1O	**50-50**	**L1O**	**50-50**	**L1O**	**50-50**
7-Moments	**79.40% (2.234)**	**80.37% (1112)**	**80.08% (2.453)**	**80.78% (2009)**	**78.90% (2.341)**	**79.97% (2.122)**
Statistics Histogram	**81.51% (1.002)**	**82.16% (1.021)**	**81.70% (1.200)**	**82.35% (0.127)**	**81.32% (1.232)**	**81.98% (1.001)**
HOG	**83.91% (3.312)**	**86.11% (1.111)**	**88.02% (3.233)**	**87.67% (1.009)**	**79.81% (3.899)**	**86.08% (1.010)**
Fractal Dimension (FD)	**85.92% (2.134)**	**86.47% (0.102)**	**86.85% (2.344)**	**86.91% (0.110)**	**84.99% (2.131)**	**86.03% (1.121)**
Gabor Filter	**87.81% (2.221)**	**89.64% (1.001)**	**88.87% (1.237)**	**90.60% (1.020)**	**86.76% (1.554)**	**88.63% (0.299)**
kNN Classifier (Model 2)						
7-Moments	**76.64% (2.112)**	**79.41% (1.231)**	**77.24% (2.243)**	**79.87% (2.121)**	**76.05% (2.888)**	**78.95% (2.211)**
Statistics Histogram	**80.01% (1.231)**	**79.99% (2.331)**	**80.03% (1.431)**	**80.76% (2.221)**	**79.99% (1.765)**	**78.56% (1.255)**
HOG	**81.20% (1.221)**	**84% (1.004)**	**81.67% (1.284)**	**84.24% (1.049)**	**80.74% (1.376)**	**83.76% (1.889)**

Fractal Dimension (FD)	**82.22% (1.343)**	**83.63% (2342)**	**83.32% (1.765)**	**84.39% (2.912)**	**82% (1.998)**	**82.88% (2.982)**
Gabor Filter	**83.28% (2.445)**	**86.10% (1.009)**	**85.36% (2.871)**	**87.56% (0.111)**	**81.62% (3.009)**	**85.65% (1.090)**

First of all, these results further confirmed that the SVM classifier is showing better results than kNN classifier for all the 5 texture features. The performances of the feature extraction techniques as shown in Table 5-3, confirm that the adaptive de-noising approach based on blocking (Model 2) improves all texture feature methods in comparisons to both no de-noising or blind de-nosing results shown in Table 5.1 above. Except for the Statistics Histogram scheme, the improvement in performance for the SVM classifier is in the range of 2-4% in absolute accuracy. The performance of the Statistics Histogram schemes may be affected, more than other schemes, as a result of significant variation in RoI size. This factor will be explored in the final section of this chapter.

Having established that the use of model 2 adaptive speckle de-noising improves the performance of each of the 7 texture-based feature schemes, it is natural to test the effect of this approach when we fuse several of these schemes at the decision level using simple majority rules. For that, we conducted a different set of experiments to evaluate the performance of several combinations of schemes (i.e. best three features, best five features and all seven features). Again, we used the SVM & kNN classifiers with both of our testing protocols. Results are shown in Table 5-4.

Table 5- 4: Performance of different Majority Voting fusion schemes based on SVM and kNN (k value =1) classifiers.

SVM Classifier						
	Accuracy mean, (Stdev)		**Sensitivity mean, (Stdev)**		**Specificity mean, (Stdev)**	
	L1O	**50-50**	**L1O**	**50-50**	**L1O**	**50-50**
Fusion of best 3 schemes	**87.25% (1.033)**	**90.37% (0.991)**	**87.64% (1.761)**	**90.76% (0.912)**	**86.87% (1.098)**	**89.99% (0.776)**
Fusion of best 5 schemes	**90.05% (1.760)**	**93.01% (0.088)**	**90.76% (1.842)**	**93.12% (1.001)**	**89.34% (1.773)**	**92.91% (0.954)**
Fusion of All 7 schemes	**88.64% (2.009)**	**89.82% (1.025)**	**89.55% (1.643)**	**90.70% (1.005)**	**87.73% (2.009)**	**88.94% (1.159)**
kNN Classifier						
Fusion of best 3 schemes	**82% (2.144)**	**86.33% (2.564)**	**83.15% (2.400)**	**87.55% (1.998)**	**80.85% (2.002)**	**85.12% (1.644)**

Fusion of best 5 schemes	**86.88% (2.200)**	**88.33% (0.876)**	**88.12% (2.641)**	**88.67% (0.981)**	**85.65% (2.198)**	**87.99% (0.999)**
Fusion of All 7 schemes	**85.81% (3.895)**	**87.70% (1.877)**	**88.52% (3.865)**	**88.77% (1.977)**	**83.11% (2.990)**	**86.64% (2.128)**

These results again confirmed that fusion at the decision level, even with simple majority rule, boasts the accuracy of texture-based diagnostic of ovarian tumour from ultrasound scan images. Moreover, there are marginal differences, if any, between Specificity and sensitivity rates for all fusion combinations. The SVM classifier leads to the best performance of over 90% for both training/testing protocols, which is comparable to the performance of well qualified medical experts.

5.4 The Impact of RoI Size Variation on the Performance of Texture-based Schemes

In this section, we shall consider the effect of variation of RoI size on the performance of certain texture-based feature schemes for diagnosing the ovarian tumour. When RoI was segmented from the experimental dataset, we noted the significant variation in their sizes. This variation possibly reflects variation in tumour size at different stages of the disease, and in the clinical setting, this problem is expected to occur frequently. Size variation has obvious impact, not only on efficiency but on the actual entries of the various image texture feature vectors and by implication on the diagnostic decisions. This effect becomes particularly adverse for texture-based features that are defined in terms of histograms when there is a significant size variation within the training images and/or at between training and testing stages. Texture features, studied in this thesis, most affected by such variation are the LBPs and the Statistics Histogram. In classification tasks, many researchers opt for normalising image size by resizing to a fixed-size image.

However, when an image is resized, its pixel information change as a result of necessary interpolation procedure which may result in reduced image quality and loss of vital information, especial if the resizing ratio is significant. In our experiments, the variation in RoI size of the images is significant, ranging from (90 x 85 to 1201 x 1100). In all previous experiments, we did not adopt any resizing policy. Figure 5-9, below, illustrate such effect on the histograms when we enlarge or reduce the size of Ultrasound ovarian tumour RoI (a) demonstrations examples about small size image with its histogram (size 120 x139) (b) resize the same image into (450 x450) with its histogram.

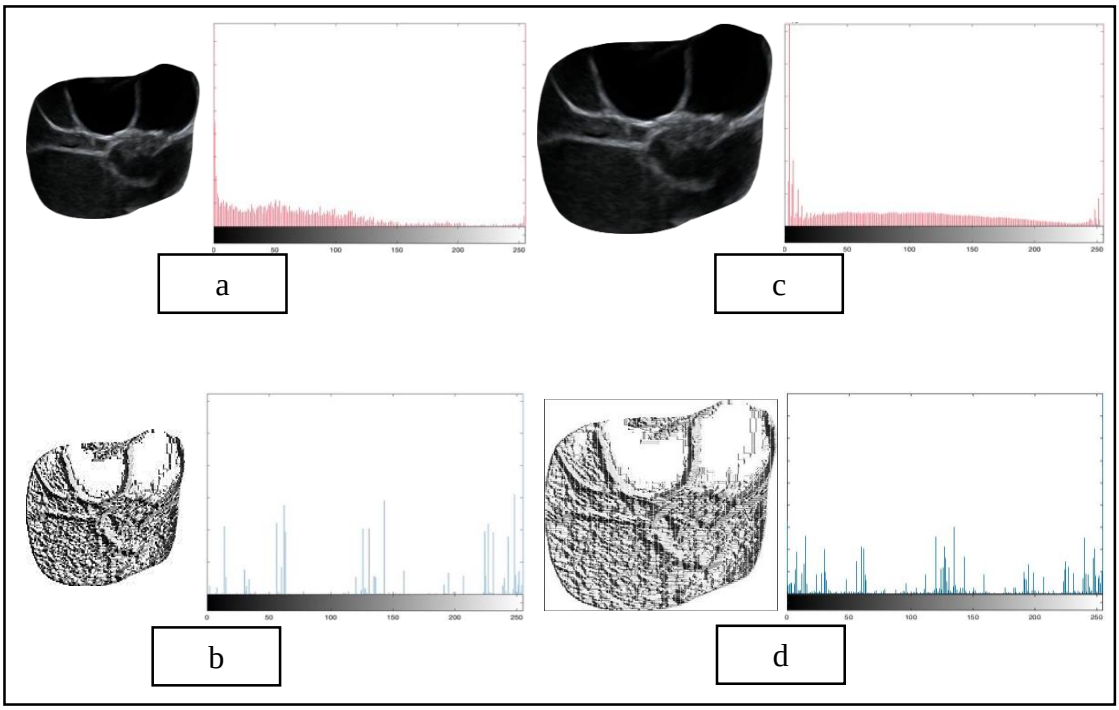


Figure 5- 9: Illustrations example of an image (a) small size of image with its histogram (b) LBP image with its histogram (c) increase the size of image with its histogram (d) LBP image with its histogram.

Another example when the image is large (1201x1100) resized into (450x450) smaller than the original dimensions, the quality of the image will affect because we will lose much information. Figure (5-10) shows an example of the impact on rescaled the image when it displays the histogram.

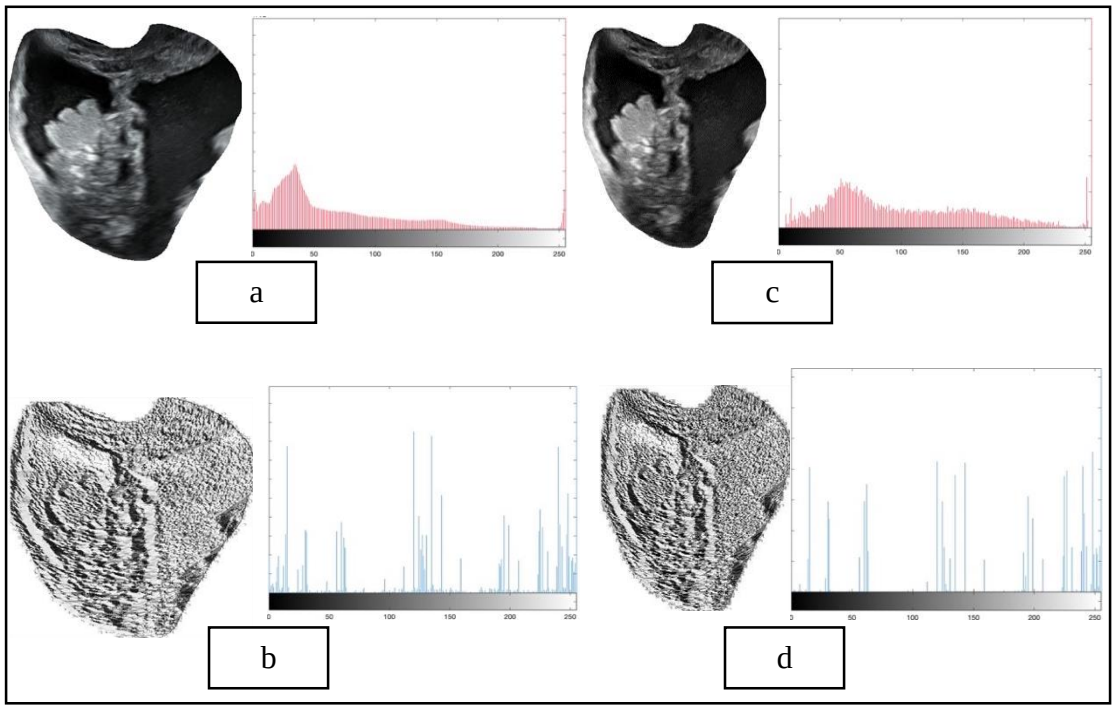


Figure 5- 10: Displays example of image (a) large size of image with its histogram (b) LBP image with its histogram (c) reduce the size of image with its histogram (d) LBP image with it is a histogram.

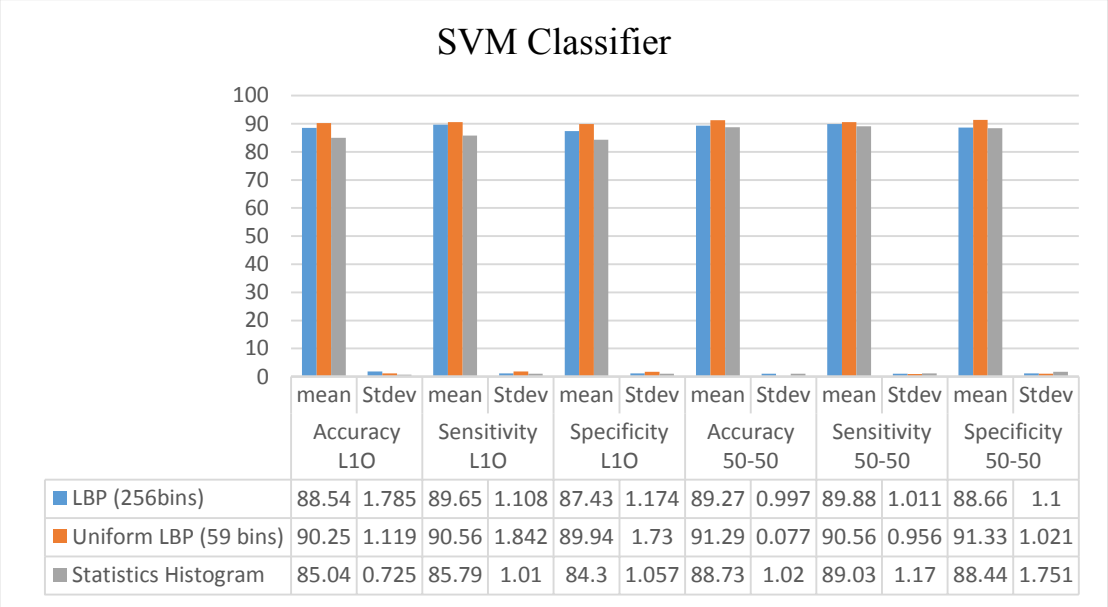
As we can see that in Figures (Figures 5.9 & 5.10), there is a clear difference between the investigated texture-based features of an input image and its resized image (i.e. the histograms, LBP histograms) due to resizing operations. Also, note the appearance of fuzzy or pixelated regions. Hence, the size normalisation process by resizing certainly has a strong influence on the feature extraction methods of (LBP 256 bins, Uniform LBP 59 bins as well as static histogram features) reducing their power of discriminating between different types of ovarian cancer. Therefore, image resizing before extract features may cause more complex and class overlapped feature space, which in turn have an adverse effect on the classifier's performances. The effect of resizing on the performance of the other texture features may not be negligible, but it is difficult to estimate.

5.5 A Proposed Size Variation Mitigating Method

Here we propose an alternative approach to image resizing, specifically applicable to the specific texture-based features discussed in this section. We propose to normalise these feature vectors rather than resizing the images prior to their extractions. This normalisation procedure is applied to each of the corresponding histogram vectors by dividing each bin value by the size of the image, i.e. if H is the feature's histogram then we normalise the histogram using the formula:

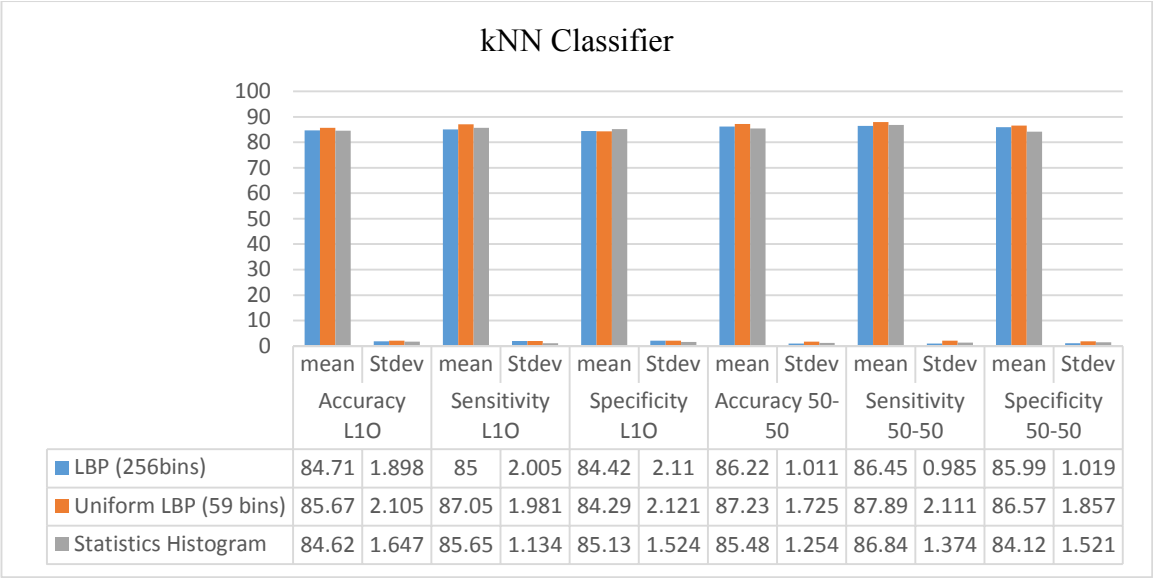
$$\text{Normalised-}H(i) = H(i) / (M * N) \tag{5.4}$$

Where M and N are the dimensions of the image, and i is the index of the bin. Then, will get new feature vectors to be fed it to classifier during both the training and testing stages. Finally, we conducted a new set of experiments to test the effect of our new idea on the performance of the three texture-based features effected by image resizing procedure: the LBP (256 bins), uniform LBP (59 bins), and the statistics histogram. As a result of the work in the previous sections, prior to this feature normalisation, we apply the adaptive pre-processing (model 2) on ROI image. Figure 5-11 & 5-12 below show the results of this new set of experiments for the SVM and kNN classifiers when trained and tested by the (L1O & 50-50) protocols.



	mean Accuracy L1O	Stdev Accuracy L1O	mean Sensitivity L1O	Stdev Sensitivity L1O	mean Specificity L1O	Stdev Specificity L1O	mean Accuracy 50-50	Stdev Accuracy 50-50	mean Sensitivity 50-50	Stdev Sensitivity 50-50	mean Specificity 50-50	Stdev Specificity 50-50
LBP (256bins)	88.54	1.785	89.65	1.108	87.43	1.174	89.27	0.997	89.88	1.011	88.66	1.1
Uniform LBP (59 bins)	90.25	1.119	90.56	1.842	89.94	1.73	91.29	0.077	90.56	0.956	91.33	1.021
Statistics Histogram	85.04	0.725	85.79	1.01	84.3	1.057	88.73	1.02	89.03	1.17	88.44	1.751

Figure 5- 11: Shows the classification performance of normalised-histogram texture-based features using SVM Classifier.



	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev
	Accuracy L1O		Sensitivity L1O		Specificity L1O		Accuracy 50-50		Sensitivity 50-50		Specificity 50-50	
LBP (256bins)	84.71	1.898	85	2.005	84.42	2.11	86.22	1.011	86.45	0.985	85.99	1.019
Uniform LBP (59 bins)	85.67	2.105	87.05	1.981	84.29	2.121	87.23	1.725	87.89	2.111	86.57	1.857
Statistics Histogram	84.62	1.647	85.65	1.134	85.13	1.524	85.48	1.254	86.84	1.374	84.12	1.521

Figure 5- 12: Illustrations the classification performance of normalised-histogram texture-based features using kNN (k value =1) classifier.

The results shown in Figure 5-11 & 5-12, demonstrate that the proposed idea of normalising the various histogram-based texture features yields significantly improved distinguishing between benign and malignant ovarian tumours with each of the three features that are defined in terms of histogram texture vectors. Overall, the increase in performance is in the range of 3-6% when compared to the results obtained in section 5.3 (Figures 5-7, 5-8 and Table 5-3).

Regarding the rest of texture-based feature schemes, used in the research such as (7-Moments, HOG, Fractal Dimension, as well as, Gabor Filter), this normalisation approach is not applicable although the negative effect of RoI resizing may not be negligible. What may be needed is more sophisticated image resizing procedures that optimally retain the texture features in the original RoI image before re-sizing such as the Super-Resolution methods (Al-Hassan 2014). This is outside the realm of this thesis and efficiency may be an obstacle.

5.6 Summary

This chapter investigated factors adversely influencing the performance of the texture-based classification scheme in medical images in an attempt to develop procedures that help to mitigate the adverse effects of these factors. In the first part of the chapter, the effect of the speckle noise on the performances of texture features was investigated. Applying Speckle noise suppression by known methods routinely on all images at training and testing was demonstrated to be at best of marginal effect. Due to the way

speckle-noise degrade different image regions that are known to have strong relevance to diagnosing in terms of presence/absence of solidity was clear guidance to develop an adaptive speckle noise suppression scheme(s). We found that training approach is far more superior to using thresholds on relevant adaptation parameter(s), and a localised block-based adaptation approach (where de-noising are applied to blocks that need it) is more relevant to our task than global adaptation (whereby the entire RoI is de-noised only if necessary). In both cases, the adaptation parameters were defined by the pair (kurtosis, skewness) of statistical parameters. In the second part of this chapter, we developed feature vector-based normalisation for texture-based diagnostic schemes that are histogram defined by histograms. We demonstrated that the proposed idea significantly increases the performances of the classification of ovarian tumours.

Chapter 6

Classification in Frequency and Transform Domains

The last two chapters were focused on developing and testing the performance of spatial domain texture-based image features for classifying Ultrasound Ovarian tumour images as benign or malignant. We also investigated the adverse impact of two factors that influence the performance of the developed schemes, namely the presence of speckle noise and significant variation in the RoI size. To mitigate the effect of these two factors, we proposed and tested a locally adaptive speckle de-noising procedure that led to significantly improved performance of each of the schemes, as well various decision level fusion of combinations of these features, and demonstrated the significant improvement in certain texture features as a result of using a feature normalisation procedure as an alternative to common approach of image resizing. Since ultrasound images, like digital camera images, can be represented equivalently by other domains such as frequency and transformed domains, then these domains expand, complement and enriches the pool of image features to be used for the classification. This chapter investigates and tests new sets of frequency and transformed domain features that are capable of discriminating the different classes. The Discrete Fourier Transform (DFT), Wavelet Transforms (including Gabor transform), and the Discrete Cosine transform (DCT) are among the most frequently deployed invertible transforms used for image processing and analysis tasks. The Local Binary Pattern (LBP) coding, and the Gradient, as well as Laplacian operators, are examples of image transforms that encapsulate a variety of image texture information.

In this chapter, we shall investigate and test the performance of several examples of texture-related features extracted from the DFT and LBP domain representations of ultrasound ovarian tumour scanned images. We shall demonstrate that these schemes have significant performance when trained and tested by the SVM and kNN classifiers on our experimental dataset.

Section 6.1 describes the image texture features using Fast Fourier Transform for ovarian tumours from US images. Section 6.2 shows the experimental results and discussion. In section 6.3, refer to the combination of different texture features based on the FFT spectrum. Section 6.4 investigates the new transformation domains using the LBP images. Section 6.5 contains a summary of this chapter.

6.1 Fourier Transform-based Image Texture Features

The spatial domain of a 2D digital image is represented by a matrix of digitized intensity values at pixels that are located at the centres of a regular rectangular grid on the scanned object. Each pixel value quantifies the aggregate of light reflected from the corresponding grid cell around the pixel. Natural light is made of a combination of waveforms carrying photons of a range of frequencies. The Fourier Transform theory is a mathematical technique, developed in the 18th century, capable of analysing the reflected light, from photographed objects, into its component frequencies in the same way as a prism analyses natural light into a rainbow. This theory is equally applicable to the analysis of another type of detectable electromagnetic radiation (or mechanically generated) waveform signals into their constituent waveform frequency range. Here, we are interested in using Fourier transform to create the frequency domain of ultrasound images for use in ovarian tumour image classification. We shall first briefly describe the mathematics of Fourier transforms and highlight the main properties of the Fourier frequency domain of images that enable effective tumour classification.

6.1.1 Fourier Spectrum of Ultrasound Images

The continuous Fourier transform (CFT) is a mapping of the infinite-dimensional vector space of certain type of complex-valued functions decomposing them into linear combinations of sine and cosine waves of different frequencies. The discrete Fourier transformation (DFT) is a version of CFT that is applicable to discrete signals such as images. The output of applying CFT/DFT on an image is a complex function defined for every pair (u,v) of frequencies real (and imaginary) values are obtained by the inner product of the entire given signal with the cosine (Respectively sine) wave functions. In another word, the Fourier function at any frequency pair is dependent on contributions from the entire image pixel values.

The discrete Fourier transformation can be calculated using the following formula (Gonzalez, Woods and Eddins 2004): Where *f* is the image, *(u,v) i*s the frequency pair, and *f (x,y)* is the spatial pixel values.

$$F(u,v)=\left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cos(2\pi\,(u\,x/M+v\,y/N)\right.$$
$$\left.-\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\sin(2\pi\,(u\,x/M+v\,y/N)\right] \qquad (6.1)$$

For computational simplicity, it is customary to represent he output complex number of the DFT is a 2-dimensional array indexed by a frequency pair (*u*,*v*):

$$[Re\big(F(u,v)\big), Im\big(F(u,v)\big)] \tag{6.2}$$

The Fourier transform is invertible, and surprisingly the computation of a function f from its Fourier transform F has an almost identical formula with minor changes.

The idea behind the use of DFT for image processing/analysis this is that the reflected frequency waveforms are dependent on the characteristics of the scanned object. The frequency waveforms that an object reflects are dependent on the textural and orientational properties of the object. In the case of ultrasound images, the textural properties are related to the geometry of the tissue components of the scanned ovary. Smooth regions in an image indicate smaller variation in grey values and this will be expressed more in the image DFT by low-frequency waveforms, but highly textured tissue regions reflect larger localised variations in pixel values and therefore have a stronger response to high-frequency waveforms. Observing and analysing such effects needs some form of visualisation.

Due to the fact that Fourier coefficients are complex numbers, it is not possible to display the output of the DFT by an easy to interpret the digital image. Moreover, displaying it as a pair of images representing the cartesian coordinates does not convey useful information about the original image. However, using the polar coordinate form of the DFT output have been known to provide an alternative displayable informative pair of images:

1. The ***DFT spectrum*** image defined by the modulus of F(u,v):

$$\| F(u,v) \| = \sqrt{(Re\big(F(u,v)\big))^2 + (IM(F(u,v)))^2} \tag{6.3}$$

2. The ***DFT phase*** image defined by the angle of F(u,v):

$$\phi\big(F(u,v)\big) = \arctan\left(\frac{\mathrm{Im}\big(\mathrm{F(u,v)}\big)}{\mathrm{Re}\big(\mathrm{F(u,v)}\big)}\right) \tag{6.4}$$

Figure 6-1 shows an ovarian ultrasound image with its FFT spectrum and phase images.

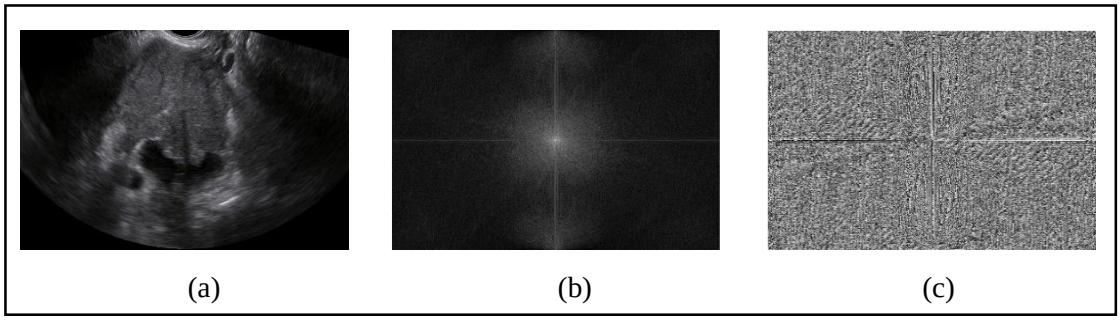


Figure 6- 1: (a) Original ultrasound ovarian tumour image (b) its DFT spectrum (c) spectrum phase.

The computation of the DFT and its spectrum and phase images are somewhat demanding. Well-known relationships between sine and cosine functions, taught at secondary school trigonometric curriculum, have been used to develop a more efficient version of DFT, called the ***Fast Fourier Transform (FFT)***. MATLAB and most programming languages provide system functions for computing FFT of an input signal/image and its inverse.

The FFT spectrum of an image has been shown to provide a good indication of the FFT's power in interpreting some aspects of the original image content. Though the Fourier spectrum of the above scan image looks featureless, however, it holds vital information on the direction of prominent features. The directions along which the spectrum is brighter correspond to the frequency orientation of image objects/texture discontinuity elements (e.g. edges and other geometric features). The discontinuities are indicated by highlighted rays that radiate from the central frequency at (0, 0) which represents the total energy (also known as the DC component in the image). Bright horizontal/vertical lines in the Fourier spectrum correspond to the image border as well as to object vertically/horizontally oriented. However, the spectrum image does not reveal any information about the position of the prominent objects/textures, i.e. it provides no spatial support. The spatial information on the image objects is visible from the phase images. Despite the above seemingly negative comments, we shall demonstrate that the FFT spectrum images do convey very useful information that enables their use for classifying Ultrasound ovarian tumour scans.

Consider the following samples of Ultrasound Ovarian Benign and Malignant tumours displayed in Figure 6-2, below, with their FFT spectrums. These images reveal interesting characteristics that we noted on a large sample of both classes of tumours that we attempt to exploit in various ways to develop frequency domain texture-based

diagnostic tools. It is not difficult to note a kind of glare around the central region for both classes, and in the case of Malignant images blurring persist over a visible size disc, but in the cases of Benign tumour the brighter lines emerge fast out of the centre are sharper and distinguishable. In fact, the central region of the FFT spectrum also seems to encapsulate vital structural information that is useful in discriminating malignant tumours from benign ones.

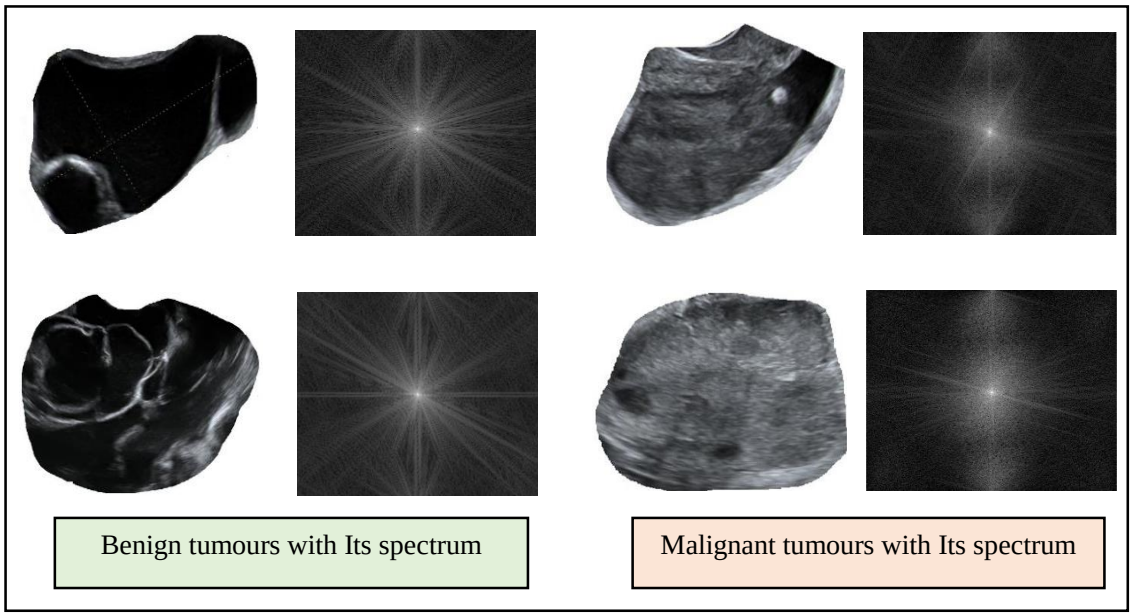


Figure 6- 2: Samples of FFT spectrum images for benign and malignant tumour RoIs.

This observation was first investigated by (Khazendar, Shan 2016), who designed the first FFT domain-based Ovarian tumour classification scheme, called the FFGF, that achieve a high accuracy rate. Next, we shall describe a modified version of Khaznadar's FFGF scheme and demonstrate improved accuracy on an expanded dataset of the one used in (Khazendar, Shan 2016) .

6.1.2 Geometric FFT based Diagnostic Schemes

The FFGF Frequency domain ultrasound ovarian image texture descriptor was designed, developed, and tested by Dr Shan Khaznadar during her PhD research program of study at Buckingham University. The proposed scheme, introduced by (Khazendar, Shan 2016) is an interesting low dimensional feature vector that can be characterised as a geometric descriptor. Our current investigation on Frequency domain texture-based Ultrasound ovarian tumour image classification was initiated first to incorporate some modifications that take advantage of the adaptive speckle de-noising scheme and to analyse the performance of the 3 components of the FFGF features separately. Later in the chapter,

we investigate an additional set of texture-features that can be extracted directly from the FFT spectrum images.

The rationale behind this that the most relevant detailed information, including the texture of the image, is represented by frequencies around the centre of the spectrum. The FFGF texture features are formed by geometric parameters associated with the centralised Fourier transform spectrum of the image. The FFGF parameters are extracted through a process of binarization via thresholding high-frequency coefficients (as indicators of significant geometric texture primitives spread throughout the spatial domain image). At a certain threshold, this process, with the help of simple morphological operations, results in a bright elliptical shape cluster, the dimensions of which form the FFGF descriptor. Figure 6-3, below, illustrates the block diagram of the FFGF of our modified version to be used for automatic classification of B-mode Ultrasound ovarian scans.

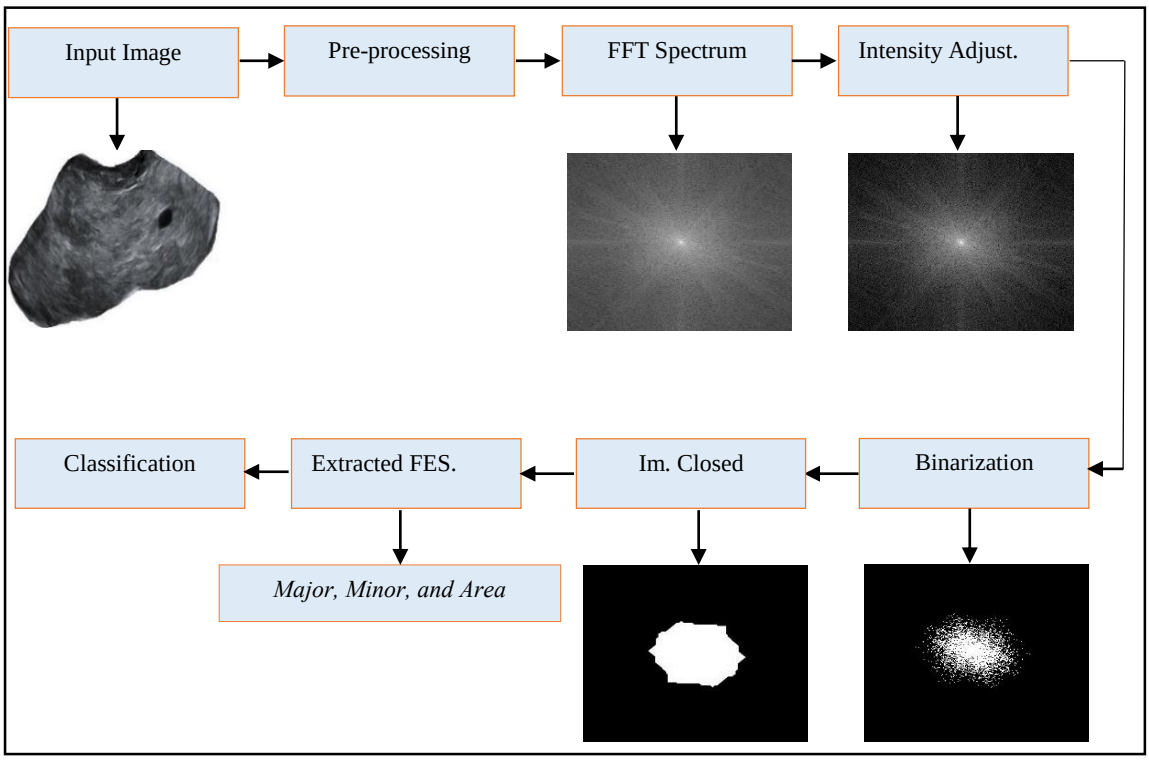


Figure 6- 3: The process of FFGF features.

The two main differences between the original and the modified FFGF schemes are:

1. Pre-processing Step – Instead of using a blanket de-noising procedure, we use adaptive block-based de-noising (i.e. model 2) as described in chapter 5. Note that in (Khazendar, Shan 2016) uses a version of non-local means based de-noising in contrast to the Wiener filter-based scheme adopted here.
2. Binarization steps – Instead of using an ad hoc intensity threshold, as practiced in (Khazendar, Shan 2016). We binarize the FFT septum image using minimum

cross-entropy to determine an elliptically shaped object in the centre followed by a morphological operation to close the ellipse. The axes and size are good indicators of the orientation of dominating textures in the original image, and others.

For more clarity, we now give a precise description of the steps of our modified fast Fourier-based geometric features (FFGF) as shown in (Figure 6.3): -

Step 1: Transforming an image into FFT frequency domain and computing the spectrum.

Step 2: Intensity adjustment

One of the image enhancement method which is used to highlight the object of the image. *i.e*. it is used to modifying the intensity values of the image by mapping the intensity values to a new range.

Step 3: Binarise the FFT power spectrum, using a minimum cross-entropy.

The cross entropy was developed by (Kullback 1997). Let F = $\{f_1, f_2, \ldots, f_N\}$ and G= $\{g_1, g_2, \ldots, g_N\}$ be two probability distributions on the same set. The cross entropy between F and G is defined by

$$D(F,G) = \sum_{i=1}^{N} f_i \log \frac{f_i}{g_i} \tag{6.5}$$

The minimum cross entropy thresholding algorithm (Li 1993) chooses a number of thresholds by minimizing the cross entropy among the original image and the resulting image. Let *I* be the original image and *h(i), i=1, 2,,L,* be the corresponding histogram with L being the number of gray levels. After that, the resulting image, denoted by I_t, based on t as the threshold value is built through

$$I_t(x,y)\begin{cases} \mu(1,t), & I(x,y) < t, \\ \mu(t,L+1), & I(x,y) \geq t, \end{cases} \tag{6.6}$$

Where

$$\mu(a,b) = \sum_{i=a}^{b-1} ih(i) \Big/ \sum_{i=a}^{b-1} h(i). \tag{6.7}$$

The cross entropy is then calculated by

$$D(t) = \sum_{i=1}^{t-1} ih(i)log\left(\frac{i}{\mu(1,t}\right) + \sum_{i=t}^{L} ih(i)\log\left(\frac{i}{\mu(t,L+1)}\right) \quad (6.8)$$

The MCET determines the optimal threshold t^* by minimizing the cross-entropy.

$$t^* = argmin_t\{D(t)\}. \quad (6.9)$$

Step 4: IMCLOSE Morphologically operation

The closing of an image is moreover a combinational operation of erosion and dilation. It is different from the opening operation in the sense of the order of occurrence of erosion and dilation operation. *A* is the closing mage which can be obtained by structuring element B that is defined in the flowing question:

$$A.B = (A \oplus B) \ominus B \quad (6.10)$$

The relation between erosion & dilation with closing is given in the above mathematical statement. It shows that closing operation is the dilation of an image *A* by the structuring element *B* and the resultant is eroded with the same structuring element. The boundary of the closed image is the points in the structuring element *B* that reaches the extreme points of the boundary of *A* when *B* is 'rolled' over A around outside of its boundary. The closing operation though smoothes sections of contours it in general blends narrow breaks and thin gaps. As a result, it eliminates small holes and fills gaps in the objects boundaries as displayed in Figure 6-4

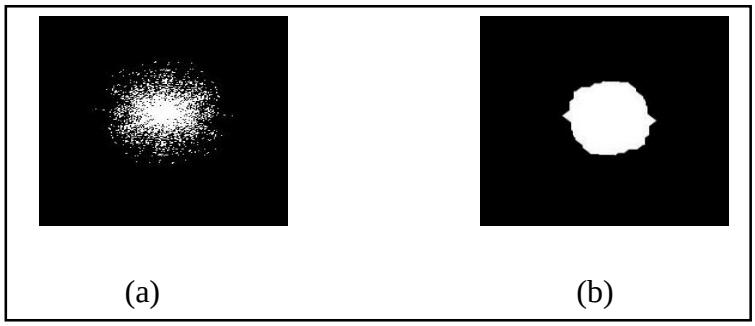


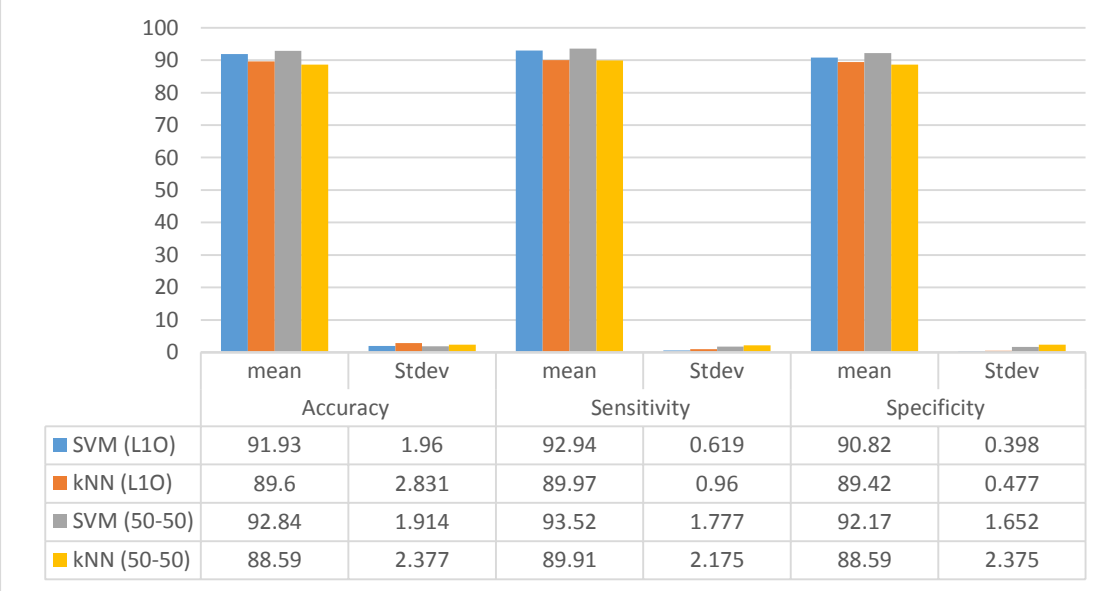
Figure 6- 4: An example of Morphologically operation: (a) Bini image (b) image after closing the objects

Step 5: 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. The three features were extracted from the fitted ellipse FFGF feature vector:

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

6.2 Experimental Results and Discussion

The experiments were conducted by training and testing SVM & kNN classifiers on our experimental dataset of 242 scan images, according to the two evaluation protocols described in Chapter 4. Figure 6-5, below, displays the results of automatic identification of the probe set of images regarding the accuracy rate and the corresponding sensitivity and specificity measures. These results demonstrate the benefits of extracting feature vectors from the FFT domain when discriminating benign from malignant tumours. More importantly, it takes only three basic frequency domain features to achieve accuracy above 89.64%. Note that the SVM classifier outperforms the kNN (where k value =1) one, despite that expectation that for low dimensional feature factors the opposite is more likely. This pattern was also notable in Khasnadar's experiments, and the obvious explanation is that the low dimensionality of the FFGF vector is dependent on every pixel in the original image as a result of the way FFT is defined.



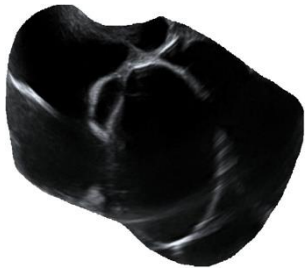
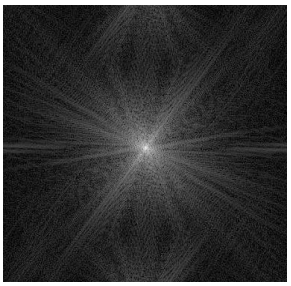
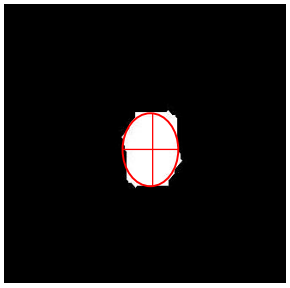
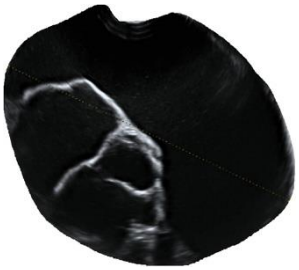
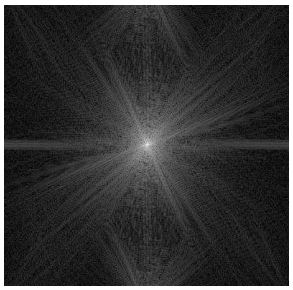
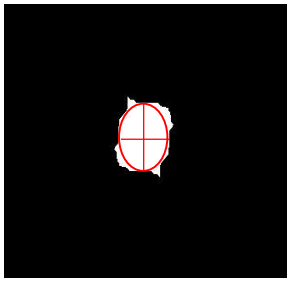
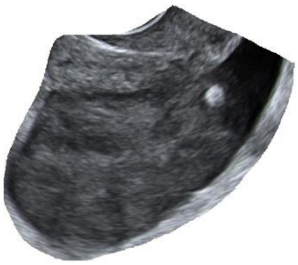
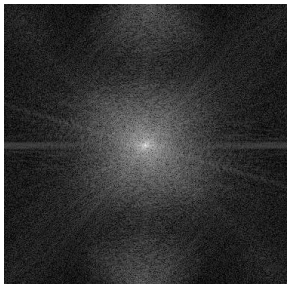
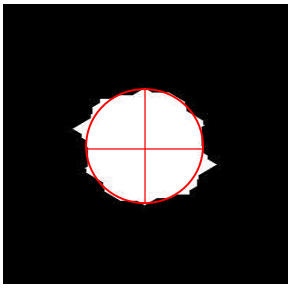
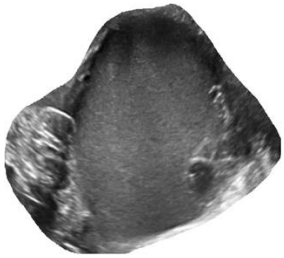
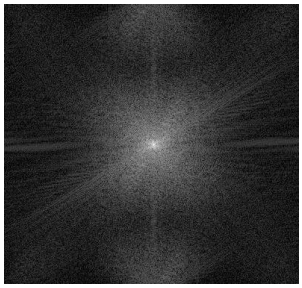
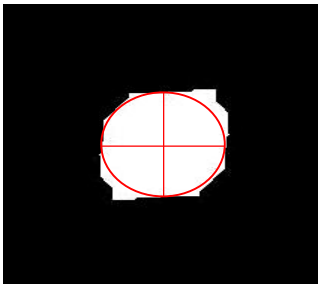
	Accuracy		Sensitivity		Specificity	
	mean	Stdev	mean	Stdev	mean	Stdev
SVM (L1O)	91.93	1.96	92.94	0.619	90.82	0.398
kNN (L1O)	89.6	2.831	89.97	0.96	89.42	0.477
SVM (50-50)	92.84	1.914	93.52	1.777	92.17	1.652
kNN (50-50)	88.59	2.377	89.91	2.175	88.59	2.375

Figure 6- 5: Classification results based on FFGF using SVM & kNN(k value =1) classifiers.

These results on the rather expanded dataset of Adnexal images that were used in demonstrate which was 85.90%; however, the modified FFGF improves significantly the performance achieved by its original version. This could be attributed to the use of the adaptive speckle de-noising in place of the blanket de-noising procedure used in (Khazendar, Shan 2016) as well as the use of entropy dependent binarization instead of the trial & error empirically determined threshold.

A close visual examination of the elliptical shapes obtained from a large sample of the FFT spectrum images of Benign and Malignant RoIs revealed an interesting variation in the shape and size of these classes. Table 6-1, below, displays a sample of benign and malignant tumour RoI together with the extracted elliptical shapes.

Table 6- 1: FFT and binary spectrums of ultrasound images of benign and malignant ovarian masses.

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

It is well known that malignant tumours tend to be larger than benign tumours due to increased solidity. In fact, malignant tumours tend to contain more complex ultrasound features, such as irregular borders, irregular echogenicity, solid parts and papillations. Interestingly, this is reflected as glares of greys near the centre of the spectrum being modelled by the extracted ellipse shape in the frequency's spectrum and the FFGF features. Malignant tumour images tend to have more glares of grey and lead to a wider minor diameter of a larger ellipse compared (Table 6-1). The above discussions, motivated us to conduct further testing on using each extracted FFGF parameter from the binary spectrum separately, i.e. Major axis, Minor axis, and area as shown in Table 6.2.

Table 6- 2: Classification result based on Minor, Major and Area using SVM & kNN based on (k value =1) classifiers.

Performance Rate / Feature Vector	Accuracy mean, (Stdev)		Sensitivity mean, (Stdev)		Specificity mean, (Stdev)	
	L1O	50-50	L1O	50-50	L1O	50-50
MinorAxisLength (SVM Classifier)	95.26 % (0.636)	95.61% (0.061)	95.75 % (0.779)	95.87% (0.651)	95.05 % (0.757)	95.35% (0.876)
MajorAxisLength (SVM Classifier)	87.80 % (1.760)	88.22 % (2.749)	88.95 % (1.282)	88.92 % (3.905)	87.20 % (1.876)	87.52 % (3.564)
Area (SVM Classifier)	89.25 % (2.913)	90.15 % (1.978)	90.73 % (3.314)	91.30 % (1.003)	87.77 % (3.175)	89.91 % (1.023)
MinorAxisLength (kNN Classifier)	94.66 % (1.221)	95.17% (2.053)	95.40 % (1.009)	95.52% (2.971)	94.12 % (1.089)	94.82% (3.918)
MajorAxisLength (kNN Classifier)	87.33% (2.541)	87.01% (01.211)	87.88% (2.111)	88.01% (1.543)	87.35% (2.652)	86.02% (1.763)
Area (kNN Classifier)	87.37% (1.121)	89.56% (1.001)	87.54% (1.765)	90.08% (0.899)	87.18% (1.101)	89.03% (1.098)

These results confirm that the minor feature alone provides the best result that even significantly outperform the accuracy reported in the previous section for the 3-dimensional FFGF feature scheme. The second-best was when the area is determined by the number of pixels in the region followed marginally by the major feature scheme. Interestingly, for each of the single parameter schemes, there are very marginal differences between the performance of the SVM and kNN classifiers. However, these significant performances, demonstrate that the FFT spectrum images of ultrasound ovarian tumour images have a very rich texture and structural information that are unfortunately obscured/hidden by its "misleading" visual representation. This observation raises a number of questions as to what other texture parameters can be extracted from the FFT spectrum and how effective such parameters are in developing additional frequency-domain diagnostic schemes for ovarian ultrasound images. Another

experiment was done by using the same process of FFGF on other type of caner which is breast cancer (see chapter 8, section 8.3).

6.3 Texture Analysis of FFT Spectrum Using Other Features

This section is devoted to answering the questions raised at the end of the last section. Taking into account, the success of the spatial domain texture schemes we developed in chapter 4 we shall extract the same 7 types of texture features from the FFT spectrum images and test their discriminating powers perhaps in comparison to the results obtained in the previous chapters. In the intended schemes, the FFT spectrum images are the source of feature vectors. Since the FFT spectrum does not provide spatial information on the objects in the original image; then we will segment the tumour RoI in the spatial domain and apply FFT on the smallest rectangular box containing the actual RoI and set the extra border pixel values to 0. Also, we should not attempt to design a special procedure to remove the effect of speckle noise on the FFT spectrum. Instead, we shall apply the adaptive speckle noise suppression procedure on the RoI prior to applying FFT.

Figure 6-6, below, describe the automated system is proposed based on the Fourier transformation. The new CAD method begins with the pre-processing (model two) then convert the pre-processed image into FFT to disapply the image spectrum. In the third stage feature extraction stage where features are identified based on texture using different feature extraction methods which are (F1) LBP256 bins, (F2) uniform (LBP), (F3) Gabor filter, (F4) fractal dimension (FD), (F5) Histograms of Oriented Gradients (HOG), (F6) Hu's invariant moments, and (F7) statistics histogram properties. Finally, the SVM & kNN classifiers are used to classify the different types of ovarian tumours.

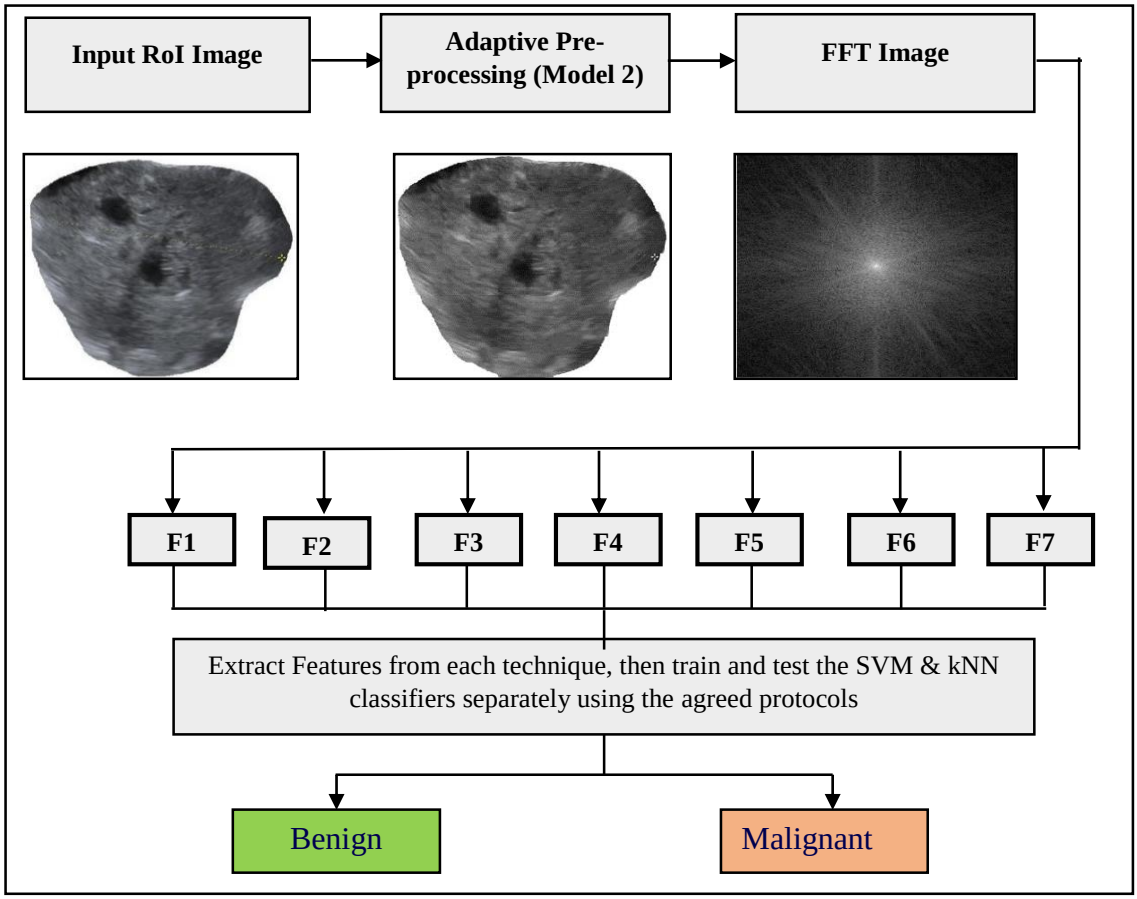


Figure 6- 6: The block diagram of FFT Spectrum texture-based classification.

In blow Figure 6-7 shows the histogram of LBP (256) based on the FFT spectrum.

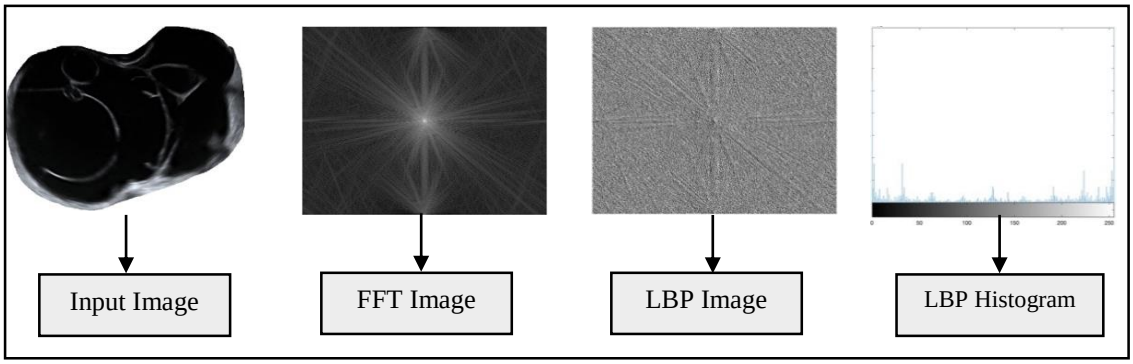


Figure 6- 7: Display example of LBP image using the FFT spectrum.

Table 6-3 displays the performances of the 7 texture schemes extracted from the FFT-spectrum when the experimental data was trained and tested by the SVM and kNN (where k value =1) classifiers according to the two agreed protocols. In these experiments, the adaptive speckle-noise reduction has been applied to the original cropped images prior to applying FFT.

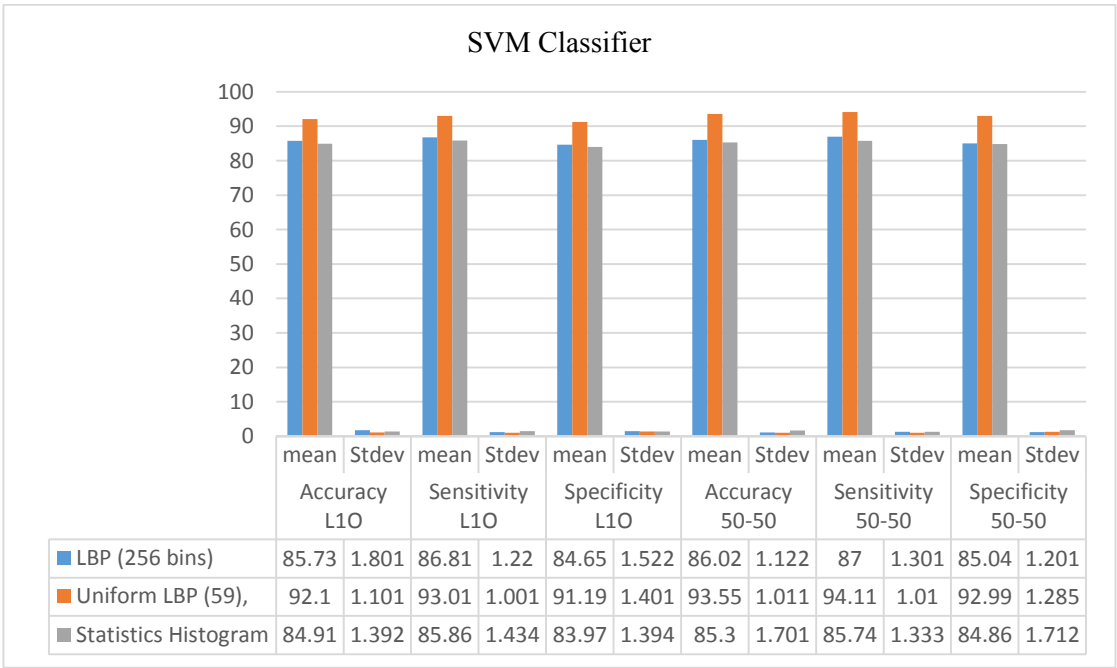
Table 6- 3: Performance of Frequency domain texture-based diagnosing schemes based on SVM and kNN (k value =1) classifiers.

SVM Classifier						
Performance Rate / Feature vector	**Accuracy mean, (Stdev)**		**Sensitivity mean, (Stdev)**		**Specificity mean, (Stdev)**	
	L1O	**50-50**	**L1O**	**50-50**	**L1O**	**50-50**
FFT_LBP (256)	**76.73% (3.125)**	**78.63% (2.912)**	**80.40% (3.096)**	**79.88% (2.711)**	**74.95% (3.275)**	**77.38% (2.003)**
FFT_LBP (59)	**88.42% (2.435)**	**89.51% (1.548)**	**91.33% (2.131)**	**90.05% (1.058)**	**87.20% (2.061)**	**88.98% (1.121)**
FFT_HOG	**85.57% (3.153)**	**85.81% (2.121)**	**88.88% (3.241)**	**87.62% (2.612)**	**83.13% (3.712)**	**84% (2.096)**
FFT_ 7 moments	**79.27% (1.431)**	**79.82% (1.008)**	**79.95% (1.193)**	**81.40% (1.201)**	**78.59% (1.622)**	**78.24% (1.101)**
FFT_ Statistics Histogram	**77.94% (1.410)**	**78.35% (2.744)**	**78.21% (0.502)**	**79.10% (2.182)**	**77.68% (1.080)**	**78.85% (2.064)**
FFT_FD	**90.22% (2.263)**	**92.64% (1.901)**	**90.82% (2.263)**	**93.87% (1.807)**	**90.46% (2.263)**	**91.40% (1.139)**
FFT_Gabor	**96.51% (1.513)**	**97.32%** (1.744)	**97.05% (1.301)**	**98.77% (1.793)**	**96.27% (1.144)**	**95.87% (1.363)**
Feature Fusion based on all Methods	**97.23% (1.044)**	**97.71% (0.949)**	**98.13% (1.101)**	**97.98% (0.711)**	**96.33% (1.114)**	**97.45% (0.783)**
kNN Classifier						
FFT_LBP (256)	**75.33% (2.753)**	**75.36% (1.720)**	**76.68% (2.043)**	**75.90% (1.827)**	**73.98% (3.623)**	**74.82% (1.993)**
FFT_LBP (59)	**75.37% (2.939)**	**78.43% (2.871)**	**77.69% (2.275)**	**79.18% (2.182)**	**73.26% (2.574)**	**77.69% (2.770)**
FFT_HOG	**77.33% (3.594)**	**80.52% (1.008)**	**80.59% (2.999)**	**81.05% (1.427)**	**76.52% (3.820)**	**80 % (1.300)**
FFT_ Hu's invariant moments	**76.47% (2.077)**	**78.34% (0.987)**	**77.87% (2.380)**	**78.64% (1.009)**	**75.08% (2.777)**	**78.03% (1.0140)**
FFT_ Statistics Histogram	**77.09% (1.998)**	**78.17% (2.587)**	**80.14% (2.287)**	**79.00% (2.199)**	**74.05% (2.089)**	**77.35% (2.731)**
FFT_FD	**84.86% (3.381)**	**85.30% (2.997)**	**88.63% (3.982)**	**87.98% (3.008)**	**81.01% (3.095)**	**82.63% (2.709)**
FFT_Gabor	**93.91% (1.713)**	**94.14% (1.101)**	**93.40% (1.981)**	**94.89% (1.098)**	**93.88% (1.729)**	**94.40% (1.243)**
Feature Fusion based on all Methods	**94.67% (1.960)**	**95.22% (1.002)**	**96.46 % (1.824)**	**96.34% (1.131)**	**92.88 % (1.919)**	**94.11% (1.129)**

These results are very interesting in that all the texture features extracted from the FFT spectrum achieved unexpectedly high accuracy, despite the fact that visually, the FFT spectrum does not seem to convey much information. These results need to be compared with the results achieved in chapter 5 when the adaptive speckle-noise reduction was used, which we are also using in these experiments. This comparison reveals that three of

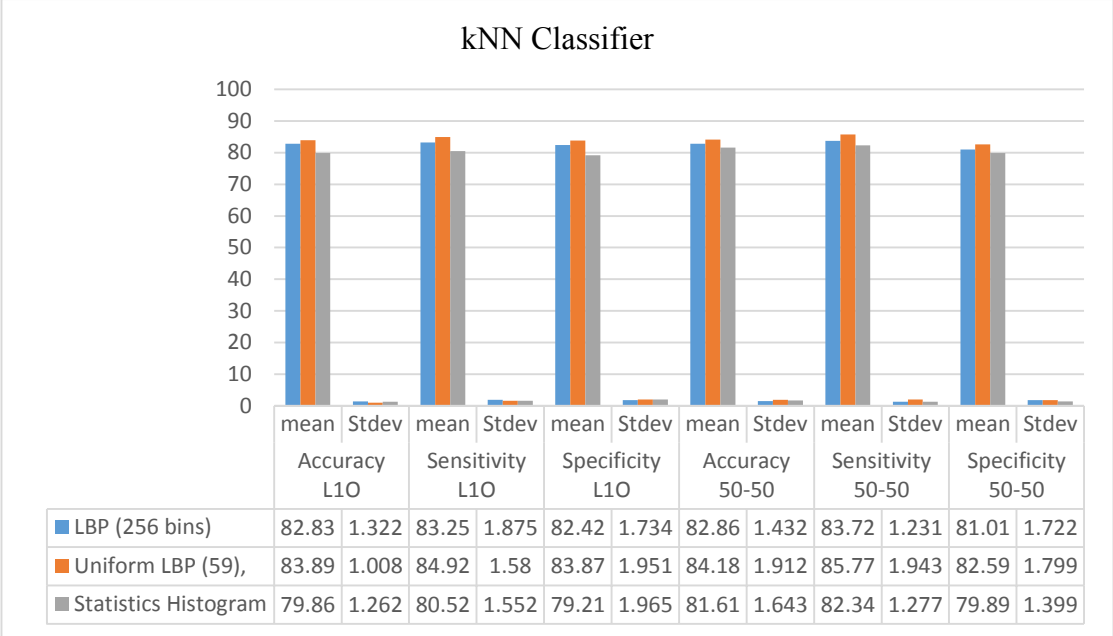
the FFT spectrum texture features (HOG, FD and Gabor) outperform their counterparts in the spatial domain and this is also true for the LBP-59 scheme but only with the SVM. The Gabor feature, in the FFT spectrum, on its own achieved the highest accuracy ever achieved by a single texture feature (>97% with SVM). Fusing all features led to a slight increase in accuracy (97.71%) compared to the single FFT-Gabor feature. The improved performance of the Fractal Dimension feature in the FFT-spectrum over its performance in the spatial domain is equally significant for both SVM and KNN classifiers. Since FD is an appropriator indicator of texture irregularity, these results show that shape irregularity is easier to detect in the frequency domain.

These results show that the other features in the FFT-spectrum compare less favourably with their counterparts in the spatial domain but the differences in accuracy are mostly marginal. One may suspect that this is due to the fact that these experiments, ignored the impact of significant variation in RoI sizes, explained in (chapter 5, section 5.5). This problem only affects (FFT-LBP-256), (FFT-ULBP (59), and (FFT -Statistics Histogram). Here we report on additional experiments to evaluate the performance of those 3 features after dividing their histogram bins by image size. Figures 6-8 & 6-9 show the performances using both protocols (L1O and 50-50) with SVM and kNN (where k value =1) classifiers, respectively.



	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev
	Accuracy L1O		Sensitivity L1O		Specificity L1O		Accuracy 50-50		Sensitivity 50-50		Specificity 50-50	
LBP (256 bins)	85.73	1.801	86.81	1.22	84.65	1.522	86.02	1.122	87	1.301	85.04	1.201
Uniform LBP (59),	92.1	1.101	93.01	1.001	91.19	1.401	93.55	1.011	94.11	1.01	92.99	1.285
Statistics Histogram	84.91	1.392	85.86	1.434	83.97	1.394	85.3	1.701	85.74	1.333	84.86	1.712

Figure 6- 8: Shows the results using SVM classifier.



	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev	mean	Stdev
	Accuracy L1O		Sensitivity L1O		Specificity L1O		Accuracy 50-50		Sensitivity 50-50		Specificity 50-50	
LBP (256 bins)	82.83	1.322	83.25	1.875	82.42	1.734	82.86	1.432	83.72	1.231	81.01	1.722
Uniform LBP (59),	83.89	1.008	84.92	1.58	83.87	1.951	84.18	1.912	85.77	1.943	82.59	1.799
Statistics Histogram	79.86	1.262	80.52	1.552	79.21	1.965	81.61	1.643	82.34	1.277	79.89	1.399

Figure 6- 9: Demonstration the results using kNN(k value =1) classifier.

First of all, these results confirm the effectiveness of our solution, for the size variation problem. Except for the LBP-59 in the FFT-spectrum, the other features remained outperformed by their spatial domain counterparts, but the gaps are more marginal. In summary, these results confirm the validity of our earlier assertion that the FFT spectrum image provides a rich source of texture-based diagnostic schemes with considerable performances. Moreover, these results provide significant opportunities for testing texture-based diagnostic schemes in other transformed images. The next section is devoted to investigating this claim for a specific image transformation.

6.4 Texture-based Diagnostic Schemes in the LBP Transform Domains

In this final section of this chapter, we test the validity of the assertion associated with the above claim, by conducting similar experiments to those conducted in the last section by replacing the FFT spectrum with the LBP image/map as a transformed domain for feature extraction. We extracted the same 7 texture features that reflected energy changes affected by specific texture and shape structures of the tumours. We shall also investigate an LBP-FFGF geometric feature by constructing the FFT spectrum of the LBP image and use the binarisation procedure in section 6.1.1 to obtain the elliptical shape and its dimensions.

In this work, the model two of pre-processing was used before applying the LBP transform. Then, different feature extraction methods which are (FFGF (**MinorAxisLength**), uniform (local binary pattern), Gabor filter, Fractal Dimension (FD), Histograms of

Oriented Gradients (HOG), Hu's invariant moments, and statistics histogram were used on LBP image to exacted texture features as shown in Figure 6-10.

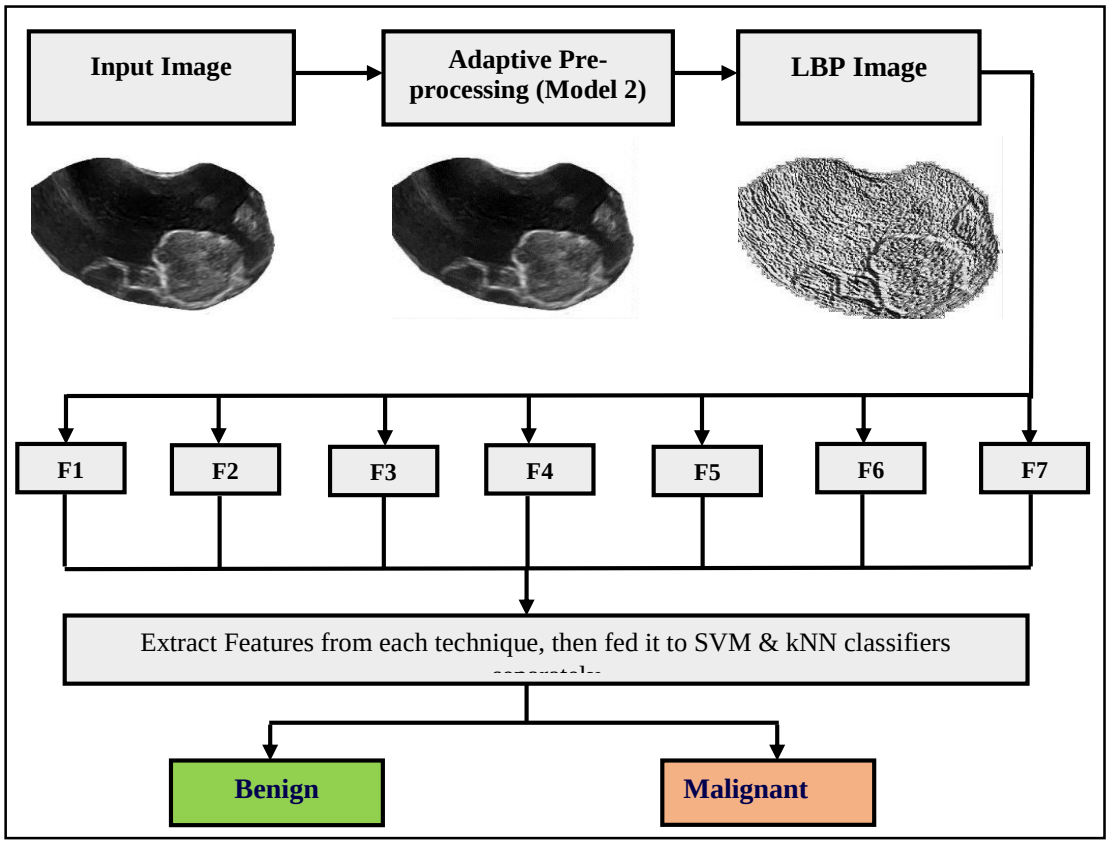


Figure 6- 10: Features extraction methods in the LBP transformed domain.

Table 6-4, below, displays the performance of each of the 7 features extracted from the LBP transformed images, as well as their fusion, using SVM & kNN (where k value =1) with (L1O & 50-50) protocols.

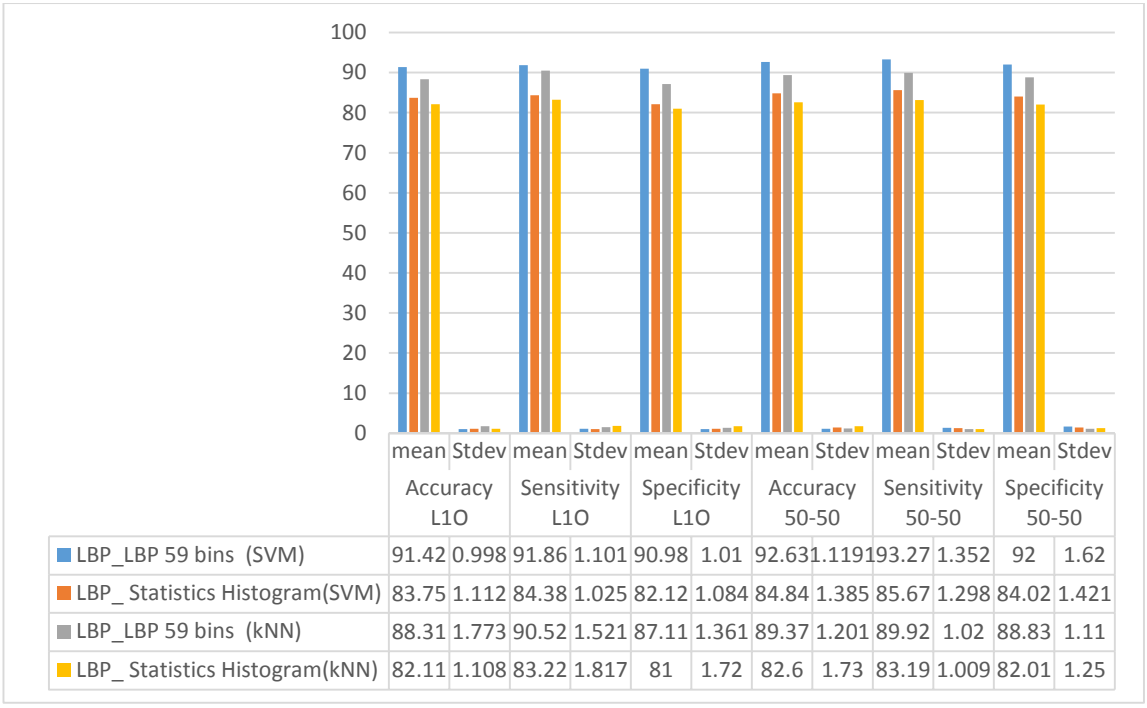
Table 6- 4: Performance of a combination of texture features extracted from LBP domain based on SVM and kNN (k value =1) classifiers.

SVM Classifier						
Performance Rate / Feature vector	Accuracy mean, (Stdev)		Sensitivity mean, (Stdev)		Specificity mean, (Stdev)	
	L1O	50-50	L1O	50-50	L1O	50-50
LBP_FFGF	93.20% (1.913)	94.56% (1.881)	93.34% (1.233)	95.22% (2.736)	92.95% (1.091)	93.89% (2.863)
LBP_LBP59 bins	89.57% (1.840)	91.57% (2.734)	91.18 % (2.322)	93.42% (2.547)	88.69 % (2.211)	89.73% (2.587)
LBP_HOG	88.75% (3.333)	88.86% (2.232)	90.93% (3.211)	89.75% (2.981)	86.82% (2.129)	87.96% (2.342)
LBP_FD	90.09% (1.232)	91.03% (0.987)	90.32% (1.221)	91.09% (1.001)	89.87% (1.876)	90.97% (0.879)
LBP_Gabor	93.23 % (1.410)	94.75% (0.341)	94.85% (0.502)	94.87% (0.765)	91.62% (1.080)	94.64% (0.654)

LBP_ Hu's invariant moments	**80.78% (2.232)**	**82% (1.651)**	**81.54 % (2.742)**	**82.13% (1.709)**	**80.02 % (2.156)**	**81.87% (1.923)**
LBP_ Statistics Histogram	**77.92% (2.232)**	**78.29% (2.323)**	**78.18% (2.232)**	**79.61% (2.143)**	**77.67% (2.232)**	**76.98% (2.098)**
Feature Fusion based on all Methods	**93.66% (2.079)**	**95.29% (1.414)**	**94.46% (2.101)**	**95.81% (1.842)**	**92.86% (2.198)**	**94.77% (1.177)**
kNN Classifier						
LBP_FFGF	**92.32% (1.121)**	**93.98% (1.066)**	**92.43% (1.192)**	**94.32% (1.341)**	**90.21% (1.121)**	**93.65% (1.753)**
LBP_LBP59 bins	**86.10% (1.543)**	**88.31% (1.001)**	**86.87% (2.098)**	**88.76% (0.998)**	**85.34% (1.987)**	**87.87% (1.098)**
LBP_HOG	**84.38% (2.987)**	**86.17% (2.112)**	**85.89% (3.009)**	**86.81% (2.876)**	**82.87% (2.787)**	**85.54% (2.016)**
LBP_FD	**87.23% (1.098)**	**87.94% (2.089)**	**87.58% (1.112)**	**88.98% (2.143)**	**86.89% (1.098)**	**86.90% (2.787)**
LBP_Gabor	**89.54% (1.343)**	**91.19% (1.239)**	**90.39% (1.987)**	**91.61% (1.125)**	**88.69% (2.101)**	**90.78% (1.143)**
LBP_ Hu's invariant moments	**76.57% (3.642)**	**78.98% (2.121)**	**79.61% (3.112)**	**80% (2.198)**	**74.22% (3.176)**	**77.87% (2.093)**
LBP_ Statistics Histogram	**76.55% (2.090)**	**77.65% (1.101)**	**77.71% (2.362)**	**78.98% (1.098)**	**76.39% (2.546)**	**78.32% (1.243)**
Feature Fusion based on all Methods	**93.49% (2.521)**	**94.53% (2.311)**	**94.98% (2.822)**	**95.19% (2.106)**	**92% (2.131)**	**93.87% (2.421)**

Again, the SVM classifier yields better performances than the kNN classifiers for each of these features, and this agrees with our previous observations. The SVM results demonstrate a similar pattern of performances to those achieved by features extracted from FFT-spectrum when compared with the performances of the spatial domain counterpart features. In relation to the top 3 performing features in the FFT-spectrum, the performance of LB_LBP-59 outperformed the FFT_LBP-59, while FFT_Gabor outperformed the LBP_Gabor and the performance of the FD features in both domains are almost identical. The LBP_FFGF (MinorAxisLength) achieves a 94.56% which is marginally lower than that in the spatial domain (see Table 6-2). In comparison, the overall accuracy is increased to around 95.61 % after fusing all seven texture features based on majority voting.

Again, the lowest-performing features in the LBP domain were the moments and the Statistics-Histogram features reaching 82% and 78%m, respectively. To determine the impact of not normalising the histograms in these features, more experiments were done to mitigate the effect of image size variation in the LBP domain for two texture feature schemes which are (LBP_LBP 59 bins & LBP_ Statistics Histogram). Figure 6-11 shows the overall accuracy, sensitivity and specificity using SVM & kNN classifiers based on (L1O & 50-50) protocols.



	mean Accuracy L1O	Stdev	mean Sensitivity L1O	Stdev	mean Specificity L1O	Stdev	mean Accuracy 50-50	Stdev	mean Sensitivity 50-50	Stdev	mean Specificity 50-50	Stdev
LBP_LBP 59 bins (SVM)	91.42	0.998	91.86	1.101	90.98	1.01	92.631	1.119	193.27	1.352	92	1.62
LBP_ Statistics Histogram(SVM)	83.75	1.112	84.38	1.025	82.12	1.084	84.84	1.385	85.67	1.298	84.02	1.421
LBP_LBP 59 bins (kNN)	88.31	1.773	90.52	1.521	87.11	1.361	89.37	1.201	89.92	1.02	88.83	1.11
LBP_ Statistics Histogram(kNN)	82.11	1.108	83.22	1.817	81	1.72	82.6	1.73	83.19	1.009	82.01	1.25

Figure 6- 11: Impact of normalising histograms in the LBP domain.

6.5 Summary

This investigations chapter was initiated by developing and testing a modification of an existing Frequency domain ultrasound ovarian tumour images to incorporate the adaptive speckle-noise scheme developed in chapter 5. Not only, the modified FFGF scheme outperformed the original scheme when tested with an expanded dataset, we have shown that single parameters of the 3-dimensional FFGF elliptical shape are sufficient to achieve significant accurate diagnosing of ovarian tumours.

Noting the significant performance of the FFGF versions extracted from the FFT spectrum images that do not easily provide spatial information on image objects led to curiosity as to investigate other frequency domain texture features that can be automatically extracted and tested for their analytical characteristics. The experimental results confirmed that indeed, the FFT frequency domain representation of the Ultrasound ovarian tumour images is a rich source of effective and highly reliable frequency texture-based tumour classification schemes. The success of these investigations provided strong motivation to search other image transform domain to further enrich the pool of texture-based diagnostic schemes. The final set of experiments on texture features extracted from the LBP transform domain demonstrated beyond any doubt the success of our approach and provided more evidence to support this stated hypothesis that image transformed domains provide a significantly larger pool than the spatial domain.

The successful outcome of these investigations indicates the huge potentiality of using machine learning to provide support for biomedical image analysis and the identification of gynaecological abnormalities.

Chapter 7

A Prospective Clinical Test of the developed schemes

The various investigations reported in the last three chapters were focused on designing effectiveness Machine Learning (ML) tools for diagnosing different types of ovarian tumours by analysing ultrasound ovarian Tumour scan images. Many different types of hand-crafted image texture-based features, in the spatial as well as transform domains, have been investigated with the aim of capturing subtle differences between benign and malignant masses. Several spatial-domain image texture-based automatic diagnostic schemes have been developed and have been shown singularly to perform well when tested with an existing dataset of ultrasound ovarian scan images labelled with the established ground truth. This research is meant to feed into the wider program in support of gynaecology clinician that aim to provide effective and reliable computational tools, for early detection of abnormalities in the ovarian, for integration into clinical diagnostic systems. The acceptability of such developed tools by the medical community is reliant on conducting clinical tests of the performance of such tools. Accordingly, the success reported in the various stages of our work motivated the launch of prospective testing of these schemes at the Department of Gynecology of Queen Charlotte's and Chelsea Hospital in London.

In section 7.1, we shall describe the clinical testing methodology, the adopted classification model, and the software interface highlighting the steps and the output format of the classification predictions. We shall also present the performance of the 7 spatial domain texture-based schemes presented in chapter 4 without pre-preprocessing, singularly and in fused combinations and determine accuracy in comparison with the gold standard histology diagnosis. The analysis of these results, in the start of section 7.2, will be used to expand the list of texture features to 9 and create a second version of the software which will also allow testing the performance of the corresponding 9 schemes with and without the adaptive pre-processing. The section ends with results of the experimental results when using with the same test cases, and demonstrate significant success for both versions in matching the gold standard histology decision. To evaluate the performance of the fused scheme and fine-tune the developed software for use in a much larger prospective testing plan with other IOTA teams in Europe. Finally, the recent rapid deployments of deep learning tools for many applications including the field of

medical diagnostics, and the various claims of significant successes was a motivation to pilot an attempt to compare the results of our prospective test with a basic implementation of a Deep CNN learning scheme trained with the original experimental dataset. This will be done in section 7.3.

7.1 Ovarian Tumor texture-based Software for Clinical Testing

A prospective IOTA 7 study was designed to test the performance of the spatial domain texture-based techniques, developed and tested in chapter 4 and chapter 5, on adnexal masses collected over a period of 6 months (starting Oct 2018) in Queen Charlotte's and Chelsea Hospital, London. Unlike the experiments, in the previous chapters no training is to be made with the B-mode static ultrasound images collected during the test, but instead, we needed to rely on the training conducted in the previous chapters. Next, we shall describe the process of selecting a trained classifier model and describe the software used in the test.

7.1.1 Model Selection for the Prospective Test

In the previous chapters, classifier training was repeated 30 times, and each time produced a different classification model. To facilitate the prospective test without using any of the test images for training a classifier, we must select a fixed classifier model (for each texture feature) to for use in making classification decision for each submitted test ultrasound ovarian tumour cropped image. We opted to choose randomly one of the classifier models obtained previously by training the SVM classifier using 242 images (104 malignant and 138 benign). Since the results of the experimental tests in the previous chapters did not reveal significant differences in performance of the various texture schemes, between the 50-50 and Leave-one-out training/testing protocols, we opted to train the classifier model using the Leave-one-out protocol.

The Randsample function in MATLAB was used to select a total of 150 images out of the 242 images (75 Benign and 75 Malignant), for training an SVM classifier for each of the 7 texture-based features. The output from this process is an SVM hyperplane model for each texture feature. Unlike the experiments in the previous chapters, we did not conduct any testing on the remaining 92 images. Therefore, the performance of the selected classifier models is not known in advance of the prospective test. When the performances were calculated, in the previous chapters, the training/testing process was repeated 30 times, and the average accuracy was shown to be significant, but with considerable standard deviation. Therefore, we do expect the performance of the chosen

models to be overfitting the features extracted from the specific 150 training images that may not be compatible with the unseen samples to be collected during the currently reported prospective test.

7.1.2 The Clinical Test Methodology

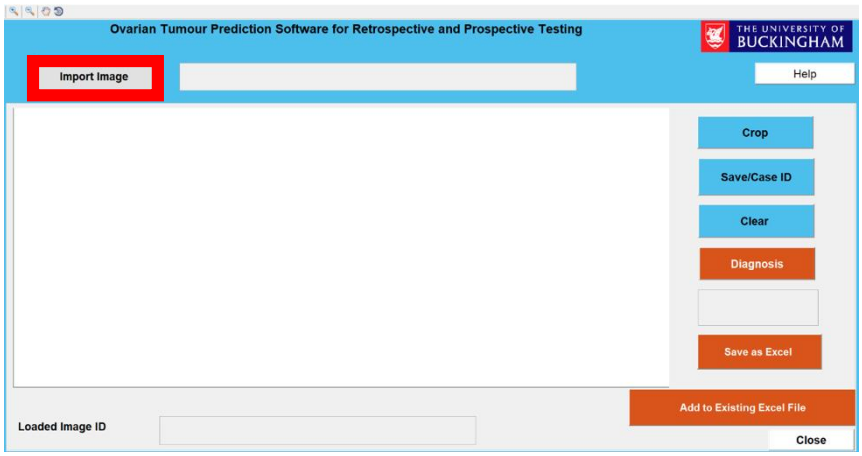
The prospective testing was conducted over the period Oct 2018 –Jan 2019. Transvaginal and transabdominal 2D B-mode static ultrasound images of 100 ovarian masses were included. All images belong to patients consecutively enrolled in the prospective IOTA 7 study from Queen Charlotte's and Chelsea Hospital in London. The entire prospective test was conducted and operated by the same highly experienced gynaecologist designated by the IOTA 7 team leaders from Leuven University Medical School and the Queen Charlotte's and Chelsea Hospital.

The most representative ultrasound ovarian scan image per test case-patient was selected by the operator (i.e. the gynaecologist who performed the ultrasound examinations). For all masses, the final histology was available for use as a standard gold test for comparison with the predicted classification by each of the texture features. The operator uploads the select image into the ML software interface, (described in details next), marks the tumour's ROI with dots and triggers the cropping of the RoI by the software. Finally, the ML software uses the chosen texture-based classifier models to output the class prediction by each texture-based scheme as well as the simple majority rule fused decision of all the schemes.

7.1.3 The Software Interface

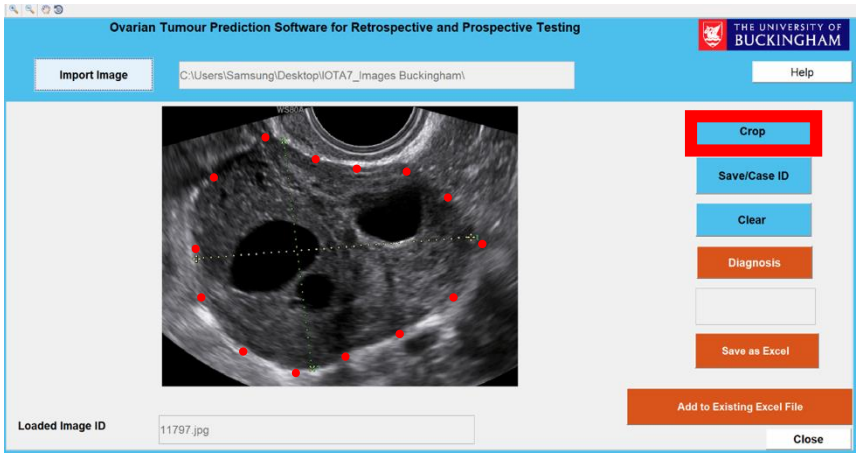
In order to facilitate the prospective test in the clinical setting without the intervention of the system developer, we provided an executable programme that was uploaded in the clinic's computer system and designed an interface for use by the clinician. The interface is meant to enable the uploading of the ultrasound ovarian scan image of the patient, currently examined for ovarian tumour abnormality, and facilitate various steps of processing analysing the input image all the way to the predicted decision. We shall now describe the main steps and facilities provided by the software. Figure 7-1 illustrations the primary interface for the software. Therefore, there are different steps to do diagnostic.

First, the expert gynaecologist will click on the import the image.



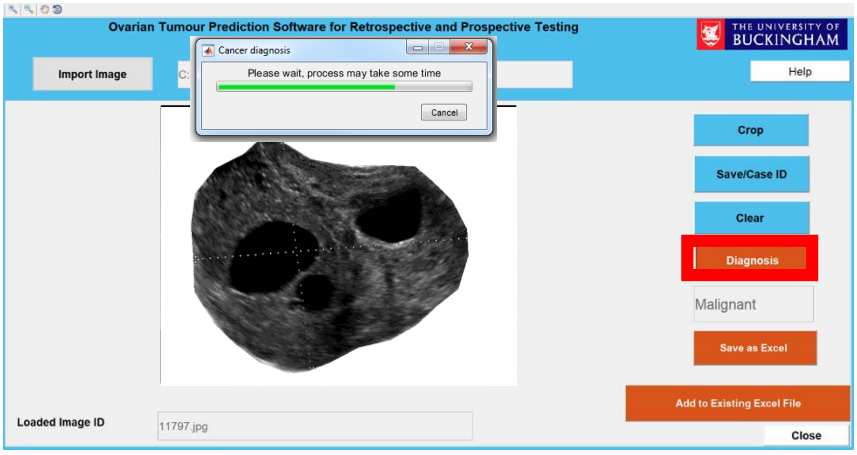
a

Send step the expert will mark the boundary of the tumour RoI and click on the crop button. The system will then crop the RoI.



b

For the software, to output, it is a class prediction, the cropped image the expert needs to click on the **Diagnosis** button, and the software displays its majority rule based fused prediction.



c

Figure 7- 1: Interface (a) import image, (b) mark RoI by expert and crop by software and (c) output a class prediction.

Moreover, the software offers the list of class predictions by all the texture schemes to be appended to an excel file to be used at the end of the prospective test for computing the performance of other fused combinations. Below is the format of this excel entry record:

Table 7- 1: Presents the format of saving decision.

0 = Benign Tumours, 1= Malignant Tumours									
Image ID	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**Final Histology**	**Machine Predication**
11799 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**Benign**	**correct**
12634 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**

7.2 Experimental Work and Prospective Test Results

The prospective test resulted in collecting B-mode static ultrasound images for 100 adnexal masses (44 malignant and 56 Benign). For each new case, the gynaecologist operator used the provided software to note the predicted class decision for the 7 spatial domain texture schemes and assessed the success/failure on the bases of majority rule fusion of all. However, the full list of predictions was saved to an excel file. Having analysed the results, and in anticipation of expanding the clinical test, a second version of the software was created first by expanding the list of schemes and then incorporating the adaptive pre-processing. We shall now present the results of testing these two versions.

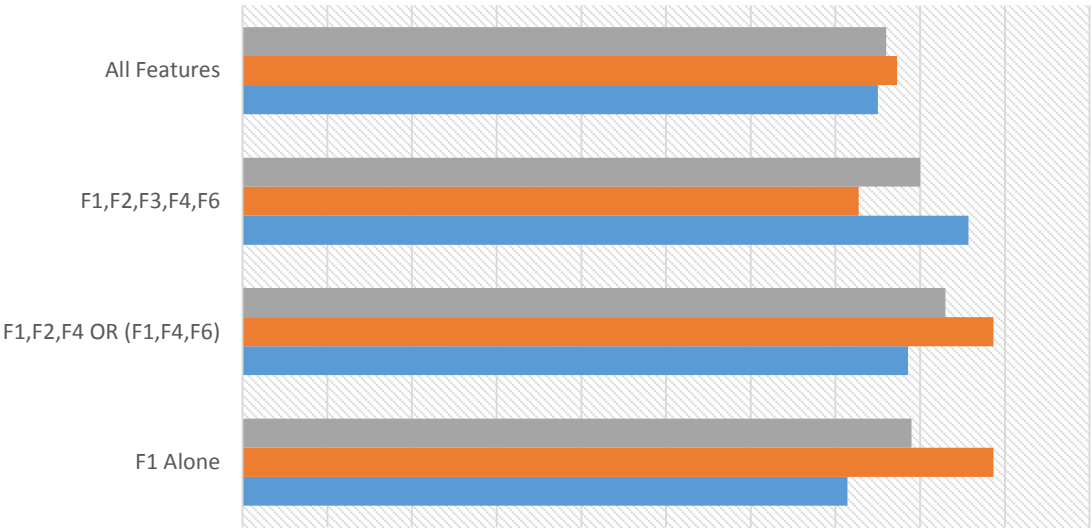
7.2.1 Performance of Version 1

Here, we present of the results of version 1 experiment by analysing the content of the used excel file (the content of which is shown in the appendix to this chapter) that recorded the list of class predictions output by the software that incorporated the 7 texture-based schemes described in chapter 4 without any pre-processing. We indexed the 7 features as follows:

F1: Gabor Filter; **F2**: Fractal Dimension; **F3:** HOG; **F4**: LBP-256 Bins; **F5**: 7-moments; **F6**: ULBP-59; **F7**: Statistics Histogram.

The excel table also contains a column showing the histology-based diagnostic decisions as well as a column that show the success/failure of the majority rule prediction. We compared several combinations of an odd number of schemes (including that of all the 7 schemes) with the histology-based diagnosis with highest confidence which are started from **F1** (in the high significant bit) and lowest confidence **F7** (in the lowest significant bit), this can reflect the general confidence of the features when converted to decimal, and this can normalise between (0 and 1) this work is ongoing. The results of these

comparisons, for some of the better performing fusion combinations, are shown in Figure 7-2 below.



	F1 Alone	F1,F2,F4 OR (F1,F4,F6)	F1,F2,F3,F4,F6	All Features
Accuracy	79%	83%	80%	76%
Sensitivity	88.64%	88.64%	72.73%	77.27%
Specificity	71.43%	78.57%	85.71%	75%

Figure 7- 2: Performance of Version 1, based on 7 schemes without pre-processing for prospective tests.

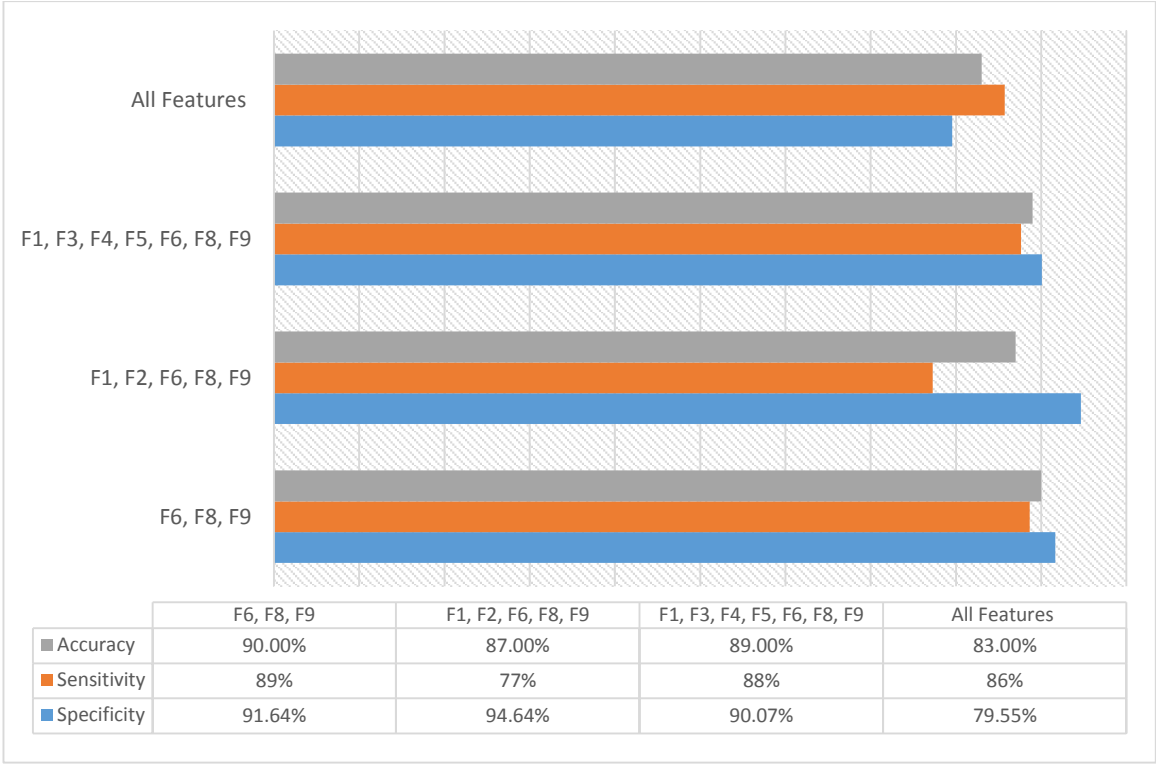
These results confirm the viability of using texture-based features for classification of ultrasound ovarian tumour images, albeit with lower accuracy than achieved in Chapter 4. The reduction in performance of the adopted schemes may reflect the fact that the randomly chosen classifier models were over fitting to the choices of the given training samples. Moreover, some of the used features involve correlated components. For example, the F5 and F7 scheme that is absent from the above list of well-performing fused combinations, include dependencies within some of their components (e.g. skewness and kurtosis). Moreover, the histograms defining the F4 and F6 share 58 of their bins. Hence, we decided to incorporate skewness and kurtosis as separate features within a 2nd version of the software.

7.2.2 Performance of Version 2

This version of the software was constructed by following the same procedures for version 1, but with the addition of two extra texture-based features, namely F8: skewness and F9: kurtosis. Motivated by the investigations carried out in chapter 5, this version of the software incorporated a choice of testing with no image pre-processing or with an adaptive speckle de-noising scheme model 2 as explained in chapter 5.

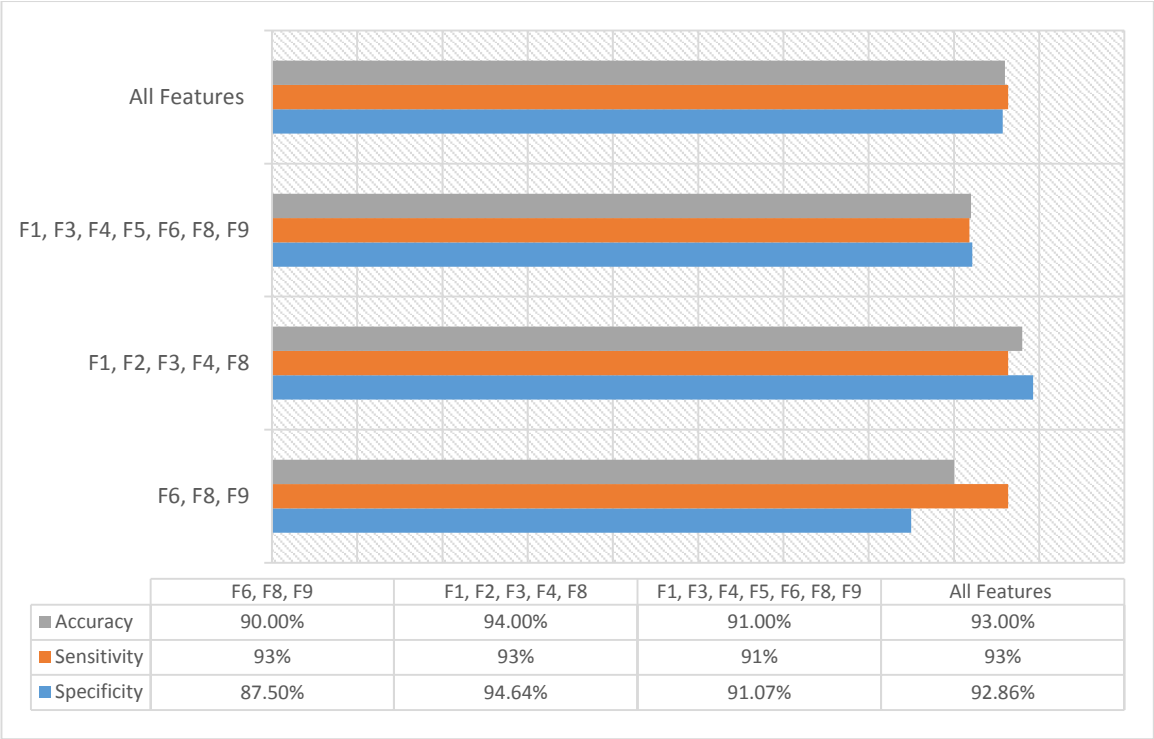
In each scenario, we repeated the running of the software for all the previously cropped tumour regions from the 100 prospective test cases described above and recorded the list of the 9 class predictions in a new excel file. The final histology, recorded earlier during the running of version 1, was used again as the gold standard for all tumours. We remind the reader that all the 9 selected classifier model were based on SVM to test the discriminating power of these 9 features (with and without pre-processing) in diagnosing Benign and Malignant ovarian tumour from Ultrasound ovary scans. Again due to the large sizes of these excel files, we display their contents in the appendix to this chapter.

As before, we compared several combinations of an odd number of schemes with the histology-based diagnosis, and the results of these comparisons, for some of the better performing fusion combinations are shown in Figure 7-3 and Figure 7-4, below, for the case of without preprocessing and with adaptive de-noising, respectively.



	F6, F8, F9	F1, F2, F6, F8, F9	F1, F3, F4, F5, F6, F8, F9	All Features
Accuracy	90.00%	87.00%	89.00%	83.00%
Sensitivity	89%	77%	88%	86%
Specificity	91.64%	94.64%	90.07%	79.55%

Figure 7- 3: Performance of Version 2, 9 schemes without pre-processing.



	F6, F8, F9	F1, F2, F3, F4, F8	F1, F3, F4, F5, F6, F8, F9	All Features
Accuracy	90.00%	94.00%	91.00%	93.00%
Sensitivity	93%	93%	91%	93%
Specificity	87.50%	94.64%	91.07%	92.86%

Figure 7- 4: Performance of Version 2, 9 schemes with adaptive de-noising.

The results of Figure 7.3 demonstrate that expanding the list with those 2 single-valued features improved the performance of the software in predicting the class of the tested tumour cases by a with the best-fused combination achieving an absolute increase in accuracy of more than 10% over the best-fused combination in version 1. Interestingly, the added 2 features (F8 and F9) appear in all the top-ranking fused combinations. Another important observations that the top two fused combinations (F6,F8,F9) and (F1,F3,F4,F5,F6,F8,F9) not only achieve almost identical accuracy rates, unlike the other combinations there are hardly any gaps between their specificity and sensitivity rates (i.e. they both similar false rejection and false acceptance).

The results shown in Figure 7.4 confirm the conclusions made in chapter 5 and demonstrate beyond any doubts the benefits in using adaptive pre-processing (i.e. adaptive speckle-noise reduction). In contrast to the previous two experiments, the fusion of all features posts adaptive pre-processing has achieved a significant accuracy improvement from 83% to 93% with almost identical false rejection and false acceptance. The fused combination (F1, F2, F3, F4, and F8) which did not even appear in Figure 7.3 achieved the highest accuracy of 94% with an equal number of false rejected and false accepted cases. The fused combination (F1, F3, F4, F5, F6, F8, and F9) did improve but only marginally, while the accuracy of the fused combination (F6, F8, and F9) suffered a marginal decline. The last pattern of accuracy changes, together with the fact that the best performing fused combination does not include F9 (i.e.

Kurtosis) may be explained by the fact that Kurtosis value is the major criteria in deciding whether an image block needs speckle de-noising or not.

The overall conclusion from these experiments shows the significant success of the prospective test which was aimed to evaluate the performance of our ML trained software that fuses several texture-based features to classify a set of unseen ovarian tumour cases in their clinical environment. In the next section, we report on a pilot study classify the prospective dataset of cases using a basic Deep Learning scheme and compare the results with the performance of our ML scheme.

7.3 Deep Learning CNN System for Ovarian Tumour Classification -A Pilot Study

Our research objectives evolved and greatly benefited from the huge advances in computer vision as well as machine learning that seem to have quickened during the course of this PhD program. ML and Artificial Intelligence (AI) both have concurrently displayed swift development over the past few years, whereby their techniques are key elements in medical field progression. Such techniques include medical image processing, computer-aided diagnosis, image interpretation, image fusion, and more. Their role in facilitating and aiding doctors for achieving disease and risk diagnosis and prediction with accuracy and speed is undeniable, alongside allowing timely disease prevention. Therefore, these techniques amplify doctors and researchers' capacity to comprehend the manner one analyses ultrasound image textures to achieve the right decision. They include standard algorithms without learning, such as Support Vector Machine (SVM), Neural Network (NN), and K-NN, as well as deep learning algorithms, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Networks (GANs), and others. Thus, this section is designed as a pilot study to investigate the use of deep learning approach in discovering wider range of image feature maps than hitherto studied in this thesis for use in classifying ultrasound ovarian tumour scan images into benign and malignant images. It is by no mean a comprehensive comparison of deep learning-based algorithms in medical image analysis issues using ML. We shall first highlight the key information and novel approaches associated with deep learning in the medical image processing and analysis landscape.

7.3.1 Deep Learning

Traditional Machine Learning (ML) models are historically built to perform beneficial tasks as per features that are manually formulated and extracted from raw data or using attributes learned by comparable simple classifier models. In the case of deep learning, computers are wired to automatically learn beneficial representations and attributes (i.e. characterising feature maps) from the raw data/images directly, with no supervision or the manual and laborious phases. At present, a multitude of artificial neural networks (ANNs) are the common models for deep learning, but there are other alternatives. Deep Learning techniques are fundamentally characterised by their emphasis on *feature learning* in which data representations are automatically learnt, in contrast to other "classical" ML techniques. Feature discovery and task performance are thus amalgamated into one problem, whereby both tasks are enhanced concurrently within the same training procedure. For an easy to follow a review of the fundamentals of the field DL, the reader is advised to refer to (oplin 2018), (Goodfellow I 2016).

Medical imaging and deep learning are linked by the common interest in *convolutional neural network* (CNN), e.g. see (Haykin 2009) which describe an excellent method of learning beneficial image representations and data structures. Prior to incorporating efficient CNN, the developer often develops a "less powerful" ML models. However, even then, the possibility of utilising attributes that are learned straight from the data results in most of the handcrafted image elements will be set aside and are rendered almost useless in comparison with CNN feature detectors. CNN's are endowed with powerful preferences according to their construction and comprehending the reasons for their classification decisions is extremely difficult, and to some extent, this is one of the main disadvantages of using DL in medical diagnostics. We shall first review the building blocks of CNN's in order to enable our pilot study.

7.3.1.1 Convolutional Neural Networks

CNN's is one of the neural network types, which is specialised as input data with a grid-like structure such as images with it. CNN's have been proven greatly effective in practical applications. Convolution is a mathematical operation which is used in at least one of its layers, instead of general matrix multiplication. CNN's are very like to regular neural networks; however, arranges its neurons in three dimensions – width, height, and depth. A neuron inside a layer is furthermore only linked to a small region of the layer before it, named the receptive field, and not fully connected as in a regular neural network. The building of CNN's involves several different kinds of sequential

layers, as shown in Figure 7-5, some of which will moreover be repeated. The most common kinds will be described below:

(i) **Convolutional Layers**: In these layers, activations from prior undergo convolvation with a set of *filters* having small parameters typically of layers 3×3 size. Each filter is precisely identical weights across the whole domain (i.e. translational equivariance at each layer) allows one to attain drastic decrement of a number of weights that require learning. Such weight-sharing is positioned by the fact that attributes present on one section of an image will also be present in other sections example for convolvation process shows in Figure 7.5. For example, the presence of a filter that can detect horizontal lines allows it to be utilised for such detection wherever they are present. Thus, the application of all convolutional filters across all input locations to a convolutional layer generates a *feature map*. (Ian Goodfellow 2016).

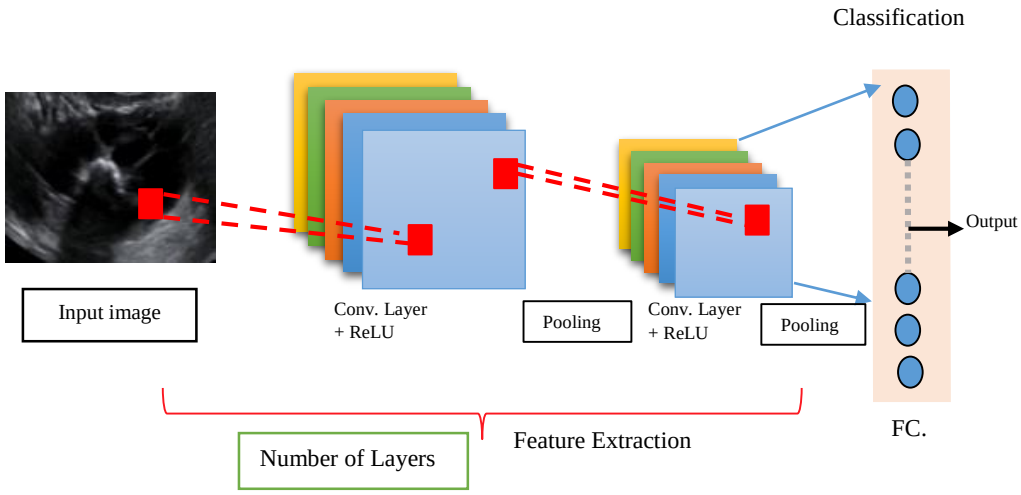


Figure 7- 5: Explanation of the general structure of a CNN

(ii) **Rectified Linear Unit Layer (ReLU):** Filters negative values to provide only positive values for much faster training time (Hijazi 2015).

(iii) **Pooling Layer (POOL):** Each feature map that is generated via data feeding using one or more convolutional layer results in them being pooled in a *pooling layer*. Such pooling operations utilise small grid regions as the input to generate single numbers for each of the regions, whereby the numbers are typically computed using the max function (*max-pooling*) or average function (*average pooling*). Even a minute shift in the input image leads to small changes in the activation maps, rendering the CNN some translational invariance. An alternative method for obtaining the downsampling impact of pooling is the incorporation of convolutions with stride length increments. Pooling layer

removal results in a simplified network arrangement without any performance being sacrificed (Springenberg 2014).

(iv) **Fully-Connected Layers:** A fully connected layer takes all neurons in the previous layer are connecting them to each of its single neurons. It is attached to a loss-function classifier (*SVM, SoftMax, Euclidean Loss* etc.) which is then used to estimate the error of the final classification and is responsible for updating the network weights through the back-propagation. (Springenberg 2014).

7.3.2 Training of CNN's

Training a CNN (Unger 2017) requires consideration of several building blocks. Which will be described now alongside the training process.

- The **score function** gets raw data as input and outputs the class probabilities. The function has a set of parameters that can be controlled; in terms of CNN's, these parameters can be trained. The goal is to adapt the weights such that the class probabilities match the ground truth labels as closely as possible. Obtaining this end calls for the **loss function,** which determines how good the prediction matches the ground truth. High loss indicates poor classification, while low loss shows good classification. The most popular loss function used with CNN's is the **cross-entropy loss** with the form:

$$L = -\sum_{i=1}^{N} (y_i \ \log q_i \ ) \tag{7.1}$$

Where q is estimated using the **SoftMax function**, the SoftMax function outputs the class probabilities between 0 and 1, which sums up to 1. The process of finding these weights that minimize the loss function is thus an optimization problem.

- The **gradient** of the loss function indicates the best direction to change the weights. **Gradient Descent** is the process of regularly performing a parameter update by calculating and evaluating the gradient. An issue associated with gradients is due to the lack of knowledge of how far one has to venture into the direction specified by the gradient. Ensuring that progress is made calls for the setting up of some step sizes, which are used to carefully follow the direction of the gradients. The step size is often referred to as the **Learning Rate**; the selection of a too-small step size yields little progress, while a too-big step size causes overshooting past the optimum. For better performance, the gradient

descent and parameter update are not performed for every training sample; it is only performed for the batches of training samples.

- **Backpropagation** is the process of calculating the gradients through the recursive use of the chain rule layer by layer of a CNN.

Before the training of a CNN, the training data need to be pre-processed in which the step consists of standardization to obtain the same range for every data sample. Additionally, as CNN's are trained for an explicit input size, the input needs to be resized or cropped to that specified size.

- Meanwhile, **Data Augmentation** is as important as the pre-processing step, as the process helps to find a model capable of providing a good generalization for the input data. It is only applied on the training set, while the test set remains unchanged except for the pre-processing. Common image augmentations include cropping, translations, rotations and flipping of input, as shown in Figure 7-6.

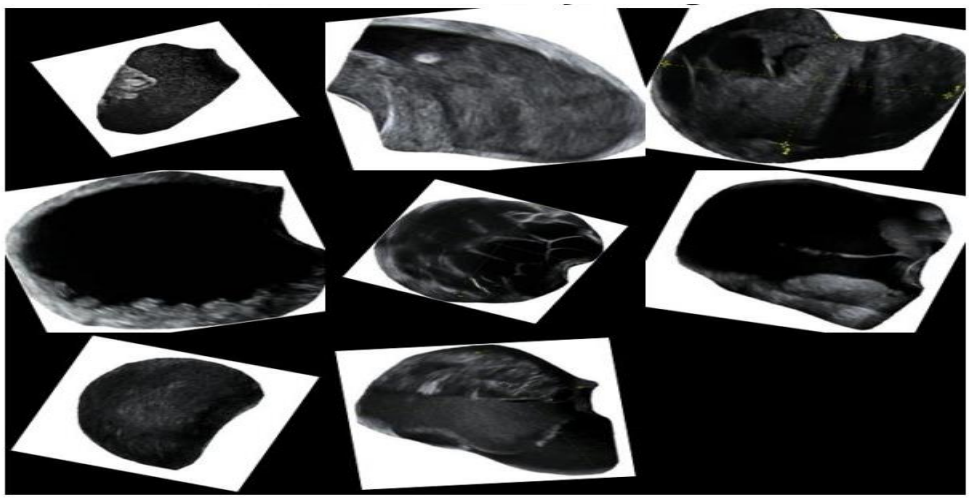


Figure 7- 6: Illustrations an example of data augmentation.

If image cropping is performed, a window of the size taken as input for each network is cut out of a bigger image and used for training.

The training process of CNN follows a pipeline. First, a batch of training samples is passed through the network to obtain class probabilities. With these probabilities and the labels of the images, the loss is calculated. Backpropagation is then used to calculate the gradient of the loss and afterwards parameter update is performed. These steps are repeated until the desired results are obtained, or no progress can be made in reducing the loss.

7.4 The Adopted CNN Architecture in This Pilot Study

A wide variety of differing CNN architectures having dissimilar characteristics are currently present. There are different CNN architectures in the literature for image classification such as ***Alexnet, VGGnet, Googlenet, Resnet50***, etc.

Resnet 50 is an example of extremely deep residual networks that was presented by He et al. (He 2016) and state of the art results was obtained on the ImageNet classification task (Russakovsky 2015). However, it has been noted that not much learning benefits are gained from using more than 30 to 50 layers since the gradient flow becomes numerically unstable in such deep networks. To alleviate the problem, a so-called residual block is presented, and layers take the form $f\hat{}\,(x) = x + f\hat{}\,'(x)$, where $f\hat{}\,'(x)$ contains the actual network layer. The main benefit in doing so is that the addition presents a second parallel branch into the network that guides the gradient flow from end to end.

ResNets50 also have other interesting properties, e.g., their residual blocks behave like ensembles of classifiers (Veit 2016), and thereby have some synergy with our approach of fusing many texture-based classification schemes for our intended application. Accordingly, our pilot will adopt the ***Resnet50*** CNN architecture.

7.4.1 Experimental Results and Discussion

Pretrained ***ResNet50*** is designed for 1000 classes; therefore, we replace the last fully connected layer with 2 output (Benign and Malignant) to output 2 posterior probabilities. For our experiments, we load weights of pre-trained CNN provided by MATLAB.

For an assessment of the performance of ***Resnet50***. A 242 images (104 Malignant, 138 Benign) tumours, then we divided the dataset into 90% for training form each class (94 images Malignant tumour, and 124 images Benign tumours) and 10% for validation (10 images Malignant tumour, and 14 images Benign tumours). Due to a small amount of dataset, we increased the number of training dataset using data augmentation by rotated the images between ($0^o - 360^0$) and augmented the training samples for every single image to 10 images, i.e. (94 images became 940 images for Malignant tumours, and 124 images became 1240 for benign tumours) with and without pre-processing each one separately. After we generated new dataset then applied ***Resnet 50***, the images should be resized to one size, which is 244 x 224 for the training and testing. We have done different

experiments with no image pre-processing or with an adaptive speckle de-noising scheme as we can see in Figures 7-7 & 7-8 the training progress after applying data augmentation.

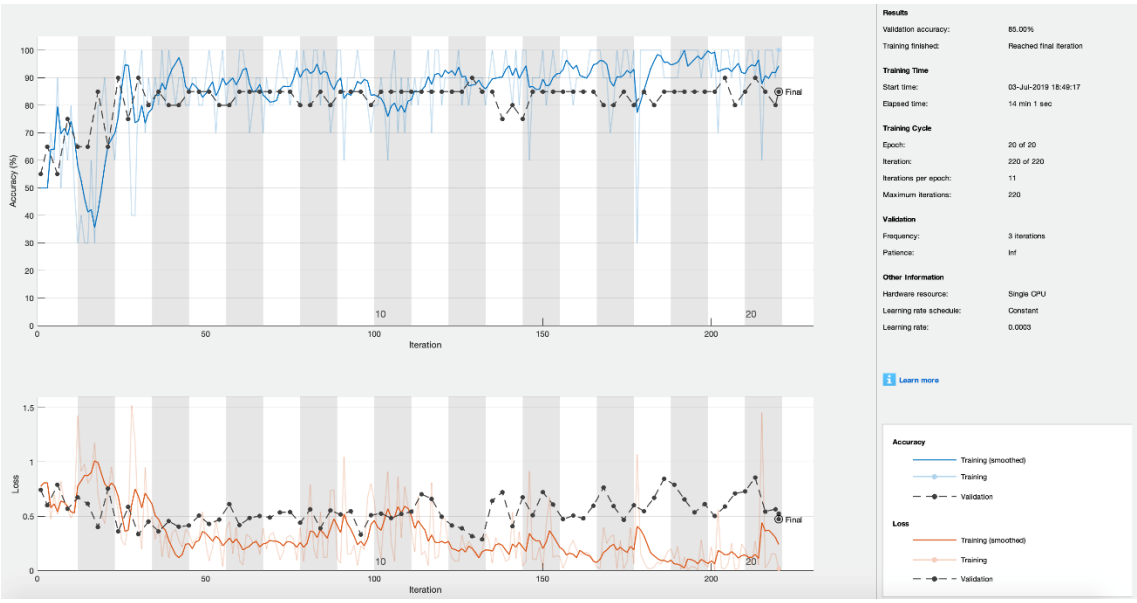


Figure 7- 7: Training progress using Resnet50 without pre-processing.

As we can see that the performance of the validation is around 85% accuracy without pre-processing scenarios. These results are matched by some of the fused combinations of the hand-craft texture-based features. Then we have trained the Resnet50 with pre-processing, which resulted in an increased validation performance reaching 95%, as shown in Figure 7.8, below.

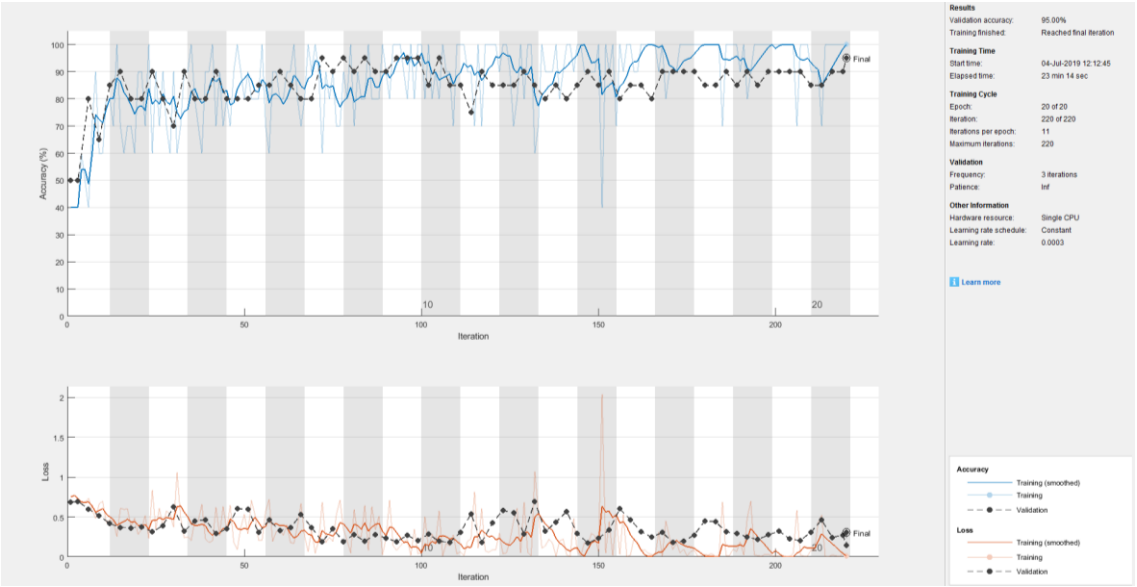
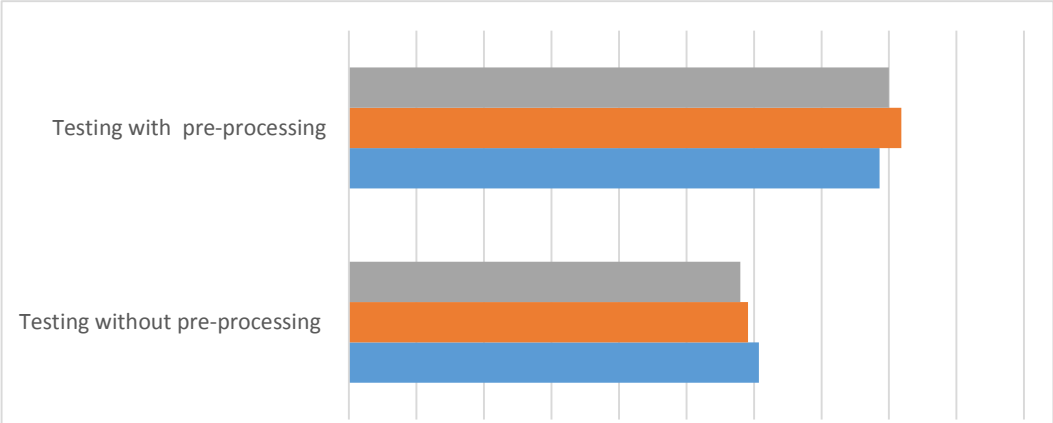


Figure 7- 8: Training progress using Resnet50 with pre-processing.

Having completed the training steps with the 242 original datasets of Ultrasound tumour images, we then tested these trained Resnet50 models, and the results are shown in Figure 7-9 for with and without pre-processing.

.



	Testing without pre-processing	Testing with pre-processing
Accuracy	79.00%	90.00%
Sensitivity	79.55%	90.91%
Specificity	80.36%	89.29%

Figure 7- 9: Testing results using Resnet50 with and without pre-processing.

Based on the above Figure 7-9 we have shown the results of this prospective clinical 100 test cases, demonstrates that the simple majority rule fusion of all the 9 features with and without the adaptive pre-processing outperform comfortably the corresponding CNN schemes. Also, as shown in Figure 7-7 and 7-8 the training progress that the clear overfitting when we tested the 100 images this due to different reasons: - Firstly the limitation of the dataset as deep learning often requires a massive dataset for the training (Camilleri 2017). Secondly, the different CNN architectures have been trained on millions of natural images. Thirdly, the effectiveness of the size of the image, i.e. the images should be one size. In the prospective test cases, images could be as large as 1200 x 1300 or as small as 90 x 85, resizing such images to 224x224 lead to great loss of information will be lost and many artefacts may appear. These results demonstrate that applying Deep learning for the classification tasks investigated in this thesis is not an effective substitute to hand-crafted features without much more in-depth investigations into different architectures. Moreover, DL is low in interpretability as it is difficult to understand what discriminating features are extracted and other limitations which are explained in (Camilleri 2017).

7.7 Summary

This chapter was designed to conclude the undertaken PhD research investigations to develop Machine Learning algorithms that use texture-based image features for automatic

classification of Ultrasound ovarian tumour scan images into benign and malignant. The research was conducted in close collaboration with leading UK and EU clinical gynaecology experts, and therefore the outcome of the research was meant to feed into wider program in support of gynaecology clinician that aim to provide useful and reliable computational tools, for early detection of abnormalities in the ovarian. The work was done over the previous chapters resulted in several high performing combinations of texture features, but integrating these schemes within clinical diagnostic systems depends heavily on the acceptability of such tools by the wider medical community via standard clinical tests of the performance of such tools. As an initial stage towards the standard clinical tests, some of the developed schemes were subjected to a prospective testing IOTA study at the Department of Gynecology of Queen Charlotte's and Chelsea Hospital in London, and in collaboration with Leuven University Medical School.

The various experiments conducted within this test confirmed the effectiveness of the various spatial domain texture-based scheme in predicting the class of the examined cases with a high level of success (well over 80% accuracy). These experiments also demonstrated that the adaptive speckle-noise reduction scheme leads to significantly improved accuracy reaching 94%.

With the recent advance in Deep learning approaches to artificial intelligence and their rapid deployments in a variety of highly complex classification problems, we conducted a pilot to test the performance of the easy to implement DL model of pre-trained Resnet50 in transfer model in classifying the prospective clinical test cases. These experiments demonstrated that this DL model did not perform as well as our schemes with or without adaptive pre-processing. Taking into account the non-informative nature of the DL decision-making process, it would be safe to conclude that it is too early to adopt DL for the classification of ovarian tumour scan images.

7.8 Appendix (Chapter 7)

Printout of Excel files showing the lists of software classification predictions for the various tests.

Table 7- 2: Display version 1 of the prospective test based on 7 Extraction Methods without Pre-processing.

Version 1. Based on (7) Feature Extraction Methods without Pre-processing									
F1: Gabor Filter; **F2:** Fractal Dimension**; F3:** HOG**; F4:** LBP-256 Bins**; F5:** 7-moments**; F6:** ULBP-59; **F7:** Statistics Histogram									
0 = Benign Tumours, 1= Malignant Tumours									
Image ID	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**Final Histology**	**Machine Predication**
11799 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**Benign**	**correct**
11935 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**Benign**	**correct**
12667 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**
12700 P.tif	**0**	**1**	**0**	**0**	**1**	**1**	**1**	**Benign**	**Not_ correct**
12777 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**Benign**	**correct**
12919 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
12920 P.tif	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**Not_ correct**
12930 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**Benign**	**Not_ correct**
12992 P.tif	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**Benign**	**Not_ correct**
13062 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**Benign**	**correct**
13063 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**Benign**	**correct**
13190 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**
13355 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**Benign**	**correct**
11972 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
13156 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**Benign**	**correct**
12958 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Benign**	**Not_ correct**
12041 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Benign**	**correct**
12165 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**
12180 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
12220 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**Benign**	**correct**
12314 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
12366 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**correct**
12378 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Benign**	**correct**
12462 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Benign**	**correct**
12492 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**correct**
12666 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**Benign**	**Not_ correct**
13356 P.tif	**0**	**1**	**0**	**0**	**1**	**1**	**0**	**Benign**	**correct**
13519 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
13520 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**
13559 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
13560 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**
13569 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**Benign**	**correct**

13601 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Benign**	**correct**
13615 P.tif	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**Benign**	**Not_ correct**
13728 P.tif	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**Benign**	**Not_ correct**
14061 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**
14098 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**correct**
14238 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**
14337 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**correct**
14380 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**Benign**	**correct**
14448 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**
14613 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**Benign**	**correct**
14621 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Benign**	**correct**
14687 P.tif	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**Benign**	**Not_ correct**
14812 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Benign**	**Not_ correct**
14813 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**Benign**	**correct**
14814 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**correct**
15039 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**Benign**	**correct**
15189 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**Benign**	**correct**
15346 P.tif	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**Benign**	**correct**
15588 P.tif	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**Benign**	**correct**
15688 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**Benign**	**Not_ correct**
15774 P.tif	**0**	**1**	**1**	**1**	**0**	**1**	**1**	**Benign**	**Not_ correct**
15773 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Benign**	**Not_ correct**
15786 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**Benign**	**correct**
14459 P.tif	**0**	**1**	**1**	**0**	**0**	**1**	**1**	**Benign**	**Not_ correct**
11797 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12699 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**Malignant**	**Not_ correct**
12831 P.tif	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12929 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Malignant**	**Not_ correct**
12954 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12969 P.tif	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**Malignant**	**Not_ correct**
12970 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
13064 P.tif	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**Malignant**	**correct**
13142 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Malignant**	**Not_ correct**
13278 P.tif	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**Malignant**	**correct**
13291 P.tif	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**Malignant**	**correct**
13308 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12015 P.tif	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**Malignant**	**correct**
12379 P.tif	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12380 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12287 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**Malignant**	**Not_ correct**
12487 P.tif	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**Malignant**	**correct**
12634 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
12646 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**Malignant**	**Not_ correct**
12665 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**

13372 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
13444 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Malignant**	**Not_ correct**
13512 P.tif	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**Malignant**	**correct**
13600 P.tif	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**Malignant**	**correct**
13611 P.tif	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**Malignant**	**correct**
13630 P.tif	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
13629 P.tif	**0**	**1**	**0**	**0**	**1**	**1**	**0**	**Malignant**	**Not_ correct**
13631 P.tif	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**Malignant**	**correct**
13750 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Malignant**	**Not_ correct**
14023 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
14024 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
14056 P.tif	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**Malignant**	**correct**
14082 P.tif	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**Malignant**	**correct**
14358 P.tif	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**Malignant**	**correct**
14359 P.tif	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**Malignant**	**correct**
14397 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
14654 P.tif	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
14733 P.tif	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**Malignant**	**correct**
14788 P.tif	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**Malignant**	**Not_ correct**
14982 P.tif	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**Malignant**	**correct**
14984 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
15431 P.tif	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**Malignant**	**correct**
15526 P.tif	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**Malignant**	**correct**
15589 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**Malignant**	**correct**

Table 7- 3: Display version 2 of the prospective test based on 9 Extraction Methods with and without Pre-processing.

Version 2. Based on (9) Feature Extraction Methods With & Without Pre-processing																						
F1: Gabor Filter; **F2:** Fractal Dimension**; F3:** HOG**; F4:** LBP-256 Bins**; F5:** 7-moments**; F6:** ULBP-59; **F7:** Statistics Histogram; **F8**: skewness; **F9**:kurtosis																						
0 = Benign Tumours, 1= Malignant Tumours																						
Without Pre-processing											**With Pre-processing**											
Image ID	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**		**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**Final Histology**	**Machine Predication (Without Pre-processing)**	**Machine Predication (With Pre-processing)**
11799 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**		**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
11935 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**Benign**	**correct**	**correct**
12667 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**	**correct**
12700 P.tif	**0**	**1**	**0**	**0**	**1**	**1**	**1**	**0**	**0**		**0**	**1**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
12777 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**Benign**	**correct**	**correct**
12919 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
12920 P.tif	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**0**		**1**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**Benign**	**Not_ correct**	**correct**
12930 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**Benign**	**correct**	**correct**
12992 P.tif	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**0**		**0**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**Benign**	**Not_ correct**	**correct**
13062 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**		**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**	**correct**
13063 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
13190 P.tif	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**		**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**Benign**	**correct**	**correct**
13355 P.tif	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**		**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
11972 P.tif	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
13156 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
12958 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**		**1**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**Benign**	**Not_ correct**	**Not_ correct**
12041 P.tif	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**		**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**Benign**	**correct**	**correct**
12165 P.tif	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**Benign**	**correct**	**correct**

12180 P.tif	0	0	0	1	1	0	1	0	0		0	0	0	1	1	0	1	0	0	Benign	correct	correct
12220 P.tif	0	0	0	0	0	0	1	0	0		0	0	0	0	0	0	0	1	0	Benign	correct	correct
12314 P.tif	0	0	0	1	1	0	1	1	0		0	0	0	1	1	0	1	1	1	Benign	correct	correct
12366 P.tif	0	0	1	0	1	0	1	0	0		0	0	0	0	1	0	0	0	0	Benign	correct	Not_ correct
12378 P.tif	0	0	0	1	0	0	1	0	0		0	0	0	1	0	0	1	0	0	Benign	correct	correct
12462 P.tif	0	0	0	1	0	0	1	0	0		0	0	0	1	0	0	0	0	0	Benign	correct	correct
12492 P.tif	0	0	1	0	1	0	1	0	0		0	0	0	0	1	0	0	0	0	Benign	correct	correct
12666 P.tif	0	0	1	1	1	0	1	0	1		0	0	1	1	1	0	1	0	1	Benign	Not_ correct	correct
13356 P.tif	0	1	0	0	1	1	0	0	0		0	1	0	0	1	1	0	0	0	Benign	correct	Not_ correct
13519 P.tif	0	0	0	1	1	0	1	0	0		0	0	0	1	1	0	0	0	0	Benign	correct	correct
13520 P.tif	0	0	0	0	0	0	0	1	0		0	0	0	0	0	0	0	1	0	Benign	correct	correct
13559 P.tif	0	0	0	1	1	0	1	0	1		0	0	0	1	1	0	1	0	1	Benign	correct	correct
13560 P.tif	0	0	0	0	1	0	0	1	0		0	0	0	0	1	0	0	1	0	Benign	correct	correct
13569 P.tif	0	0	1	0	1	0	0	0	1		0	0	1	0	1	0	0	0	1	Benign	correct	correct
13601 P.tif	0	0	0	1	0	0	1	0	0		0	0	0	1	0	0	1	0	0	Benign	correct	correct
13615 P.tif	0	1	1	0	1	1	1	0	0		0	0	0	0	1	0	0	0	0	Benign	Not_ correct	correct
13728 P.tif	0	1	0	1	0	1	1	0	0		0	0	0	0	0	0	1	0	0	Benign	correct	correct
14061 P.tif	0	0	0	0	1	0	0	0	0		0	0	0	0	1	0	0	0	0	Benign	correct	correct
14098 P.tif	0	0	1	0	1	0	1	1	0		0	0	1	0	1	0	1	1	0	Benign	correct	correct
14238 P.tif	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	Benign	correct	correct
14337 P.tif	0	0	0	1	1	0	1	1	0		0	0	0	1	1	0	0	1	0	Benign	correct	correct
14380 P.tif	0	0	1	0	1	0	0	0	0		0	0	1	0	1	0	0	0	0	Benign	correct	correct
14448 P.tif	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	Benign	correct	correct
14613 P.tif	0	0	1	0	1	0	0	1	0		0	0	1	0	1	0	0	1	0	Benign	correct	correct
14621 P.tif	0	0	0	1	0	0	1	0	0		0	0	0	1	0	0	0	0	0	Benign	correct	correct

14687 P.tif	1	0	0	1	1	0	1	0	0		1	0	0	1	1	0	1	0	0	Benign	correct	correct
14812 P.tif	0	1	1	1	1	1	1	0	1		0	0	1	1	0	1	1	1	1	Benign	Not_ correct	correct
14813 P.tif	0	0	0	0	1	0	1	0	0		0	0	0	0	1	0	1	0	0	Benign	correct	Not_ correct
14814 P.tif	0	0	1	0	1	0	1	0	1		0	0	1	0	0	0	0	0	1	Benign	correct	correct
15039 P.tif	0	0	0	0	1	0	1	1	0		0	0	0	0	1	0	1	1	0	Benign	correct	correct
15189 P.tif	0	0	1	0	1	0	1	0	1		0	0	0	0	0	0	0	0	1	Benign	correct	correct
15346 P.tif	0	0	1	1	0	0	0	0	0		0	0	0	1	0	0	0	0	0	Benign	correct	correct
15588 P.tif	1	0	1	0	0	0	1	0	0		1	0	1	0	0	0	1	0	0	Benign	correct	correct
15688 P.tif	0	0	1	1	1	0	1	0	0		0	0	0	1	1	0	1	0	0	Benign	correct	correct
15774 P.tif	0	1	1	1	0	1	1	0	0		0	1	1	0	0	1	0	0	0	Benign	Not_ correct	correct
15773 P.tif	0	1	1	1	1	1	1	0	1		0	0	1	0	0	0	0	0	1	Benign	Not_ correct	correct
15786 P.tif	0	0	0	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	Benign	correct	correct
14459 P.tif	0	1	1	0	0	1	1	0	0		0	0	0	0	0	1	0	1	1	Benign	correct	correct
11797 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
12699 P.tif	0	0	0	0	0	0	1	0	0		0	1	1	1	1	0	1	0	0	Malignant	Not_ correct	correct
12831 P.tif	1	1	0	1	1	1	1	1	1		1	1	0	1	1	1	1	1	1	Malignant	correct	correct
12929 P.tif	0	0	0	1	0	0	1	1	1		0	1	1	1	1	0	1	1	1	Malignant	Not_ correct	correct
12954 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
12969 P.tif	0	0	1	0	1	0	0	1	1		1	1	1	0	1	0	0	1	1	Malignant	Not_ correct	correct
12970 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
13064 P.tif	1	0	0	1	1	0	1	1	1		1	1	1	1	1	0	1	1	1	Malignant	correct	correct
13142 P.tif	0	0	0	1	0	0	1	1	0		0	1	1	1	1	1	1	1	0	Malignant	Not_ correct	correct
13278 P.tif	1	0	1	1	1	0	1	1	1		1	1	1	1	1	0	1	1	1	Malignant	correct	correct
13291 P.tif	0	1	1	0	1	1	1	1	0		0	1	1	0	1	1	1	1	0	Malignant	correct	correct
13308 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct

12015 P.tif	1	1	1	1	0	1	1	0	1		1	1	1	1	0	1	1	0	1	Malignant	correct	correct
12379 P.tif	1	1	0	1	1	1	1	1	0		1	1	0	1	1	1	1	1	0	Malignant	correct	correct
12380 P.tif	1	1	1	1	1	1	1	0	1		1	1	1	1	1	1	1	0	1	Malignant	correct	correct
12287 P.tif	0	0	0	1	0	0	1	1	1		0	1	1	1	0	1	1	1	1	Malignant	Not_ correct	correct
12487 P.tif	1	1	0	0	1	1	1	1	1		1	1	0	0	1	1	1	1	1	Malignant	correct	correct
12634 P.tif	0	1	1	1	1	1	1	1	1		0	1	1	1	1	1	1	1	1	Malignant	correct	correct
12646 P.tif	0	0	0	1	1	0	0	1	1		1	0	1	1	1	1	0	1	1	Malignant	Not_ correct	correct
12665 P.tif	0	1	1	1	1	1	1	1	1		0	1	1	1	1	1	1	1	1	Malignant	correct	correct
13372 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
13444 P.tif	0	0	0	0	0	0	0	0	0		0	0	0	1	0	0	0	0	0	Malignant	Not_ correct	Not_ correct
13512 P.tif	1	1	1	1	0	1	1	1	0		1	1	1	1	0	1	1	1	0	Malignant	correct	correct
13600 P.tif	1	0	1	1	1	0	1	1	1		1	0	1	1	1	0	1	1	1	Malignant	correct	correct
13611 P.tif	1	0	1	1	1	0	1	1	1		1	0	1	1	1	0	1	1	1	Malignant	correct	correct
13630 P.tif	0	1	0	1	1	1	1	1	1		0	1	1	1	1	1	1	1	1	Malignant	correct	correct
13629 P.tif	0	1	0	0	1	1	0	1	0		0	1	1	0	1	1	0	1	1	Malignant	Not_ correct	correct
13631 P.tif	0	1	0	1	0	1	1	1	1		0	1	0	1	0	1	1	1	1	Malignant	correct	correct
13750 P.tif	0	0	0	0	0	0	0	0	0		0	0	0	0	1	0	0	0	0	Malignant	Not_ correct	Not_ correct
14023 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
14024 P.tif	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
14056 P.tif	0	1	1	0	1	1	1	0	1		1	1	1	0	1	1	1	0	1	Malignant	correct	correct
14082 P.tif	1	0	1	1	0	0	1	1	1		1	1	1	1	0	0	1	1	1	Malignant	correct	correct
14358 P.tif	1	0	0	1	1	0	1	0	1		1	0	0	1	1	0	1	0	1	Malignant	correct	correct
14359 P.tif	1	1	0	0	1	1	1	1	1		1	1	1	1	1	1	1	1	1	Malignant	correct	correct
14397 P.tif	0	1	1	1	1	1	1	1	0		1	1	1	1	1	1	1	1	0	Malignant	correct	correct
14654 P.tif	0	1	1	1	1	1	1	1	1		0	1	1	1	1	1	1	1	1	Malignant	correct	correct

14733 P.tif	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**1**	**1**		**1**	**0**	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**Malignant**	**correct**	**correct**
14788 P.tif	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**1**	**1**		**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**	**correct**
14982 P.tif	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**		**0**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**	**correct**
14984 P.tif	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**		**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**	**correct**
15431 P.tif	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**		**0**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**Malignant**	**correct**	**correct**
15526 P.tif	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**0**		**0**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**0**	**Malignant**	**correct**	**Not_ correct**
15589 P.tif	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**1**	**1**		**0**	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**Malignant**	**correct**	**correct**

Chapter 8

Conclusion and Future work

Accurate and early diagnosis is between the crucial factors in the management of diseases and more so for cancer as it will determine kind of treatment in addition to prognosis. In fact, the main challenges in treating cancer patients are the fact that in most cases, the cancerous tumour is detected at advance stages of the disease when the only option available is the managing end of life. Ovarian cancer is a typical example of a disease where most cases are unfortunately diagnosed by chance and often too late for prolonged survival. This thesis was designed to contribute to the efforts of developing reliable ovarian tumour classification tools that on the one hand exploit advances in Machine learning and computer vision and on the other hand are sufficiently informative to be integrated within clinical setup.

Ovarian cancer blood test (CA-125) is usually used in conjunction with radiological results for the prediction of malignancy. Between the different kind of imaging modalities available to the radiologist, the ultrasound modality and in particular B-mode imaging is becoming the main imaging procedure for the triage of ovarian cancer. Despite its proven usefulness, the main issue with the manual examination is that it is operator-dependent, and thus the accuracy and reproducibility of the diagnosis very much subject to the experience of the operator. The shortage of highly trained sonographers and gynaecologists exasperates this problem and overburden healthcare systems throughout the world. Automating the process of examining and applying advanced image analysis of huge numbers of ovarian tumour scan images generated daily is a rather urgent task. This thesis is meant to contribute to this noble effort by complementing 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 is not directly observable/accessible from the image without applying sophisticated mathematical transformations. Our work will be confined to greyscale (B-mode) ultrasound image modality for classification of ovarian tumours, with a view to providing software tool(s) that is able to assist clinicians in their diagnosis of ovarian cancer. However, many of the results and conclusions are expected to provide useful insights for other biomedical image modalities and for other types of tumours.

This chapter highlights the main findings of the research conducted in each of the identified tasks and draws conclusions and implications for research beyond the stated objectives. Future research directions will also be outlined, with a focus on enhancing the performance and reliability of the proposed systems in analysing the ovarian ultrasound images.

8.1 Conclusion

Ultrasound imaging is considered as one of the most important and effective imaging modalities in detecting gynaecological abnormalities; however analysing these images is highly dependent on operator experience, which is in short supply in most health services. Moreover, the 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. Automating parts of this process are thus essential for more effective healthcare systems, but for variety reasons, such systems should never aim to be a substitute for training more clinicians without whom the current research could not succeed.

Having reviewed existing research in this field and considered the frequent discussions with our medical collaborators, it became clear that sonographers and gynaecologists try to detect changes to certain textural and structural image features as indicators of malignancy. This learning journey led to a kind of hypothesis on the direction of research which can be summarised as: "in order to reduce the operator-dependency, we need to adapt image texture analysis approaches and investigate several known image texture primitives and features with a focus on quantifying their tumour discriminating characteristics". This hypothesis is also supported by an evolving systems biology assumption of the way carcinogenesis results in changing the texture of cysts cellular networks.

Image texture analysis has been widely investigated in many computer vision applications, including satellite photography, biometric identification, face recognition, handwritten text analysis and document processing. The literature on such applications provides ample examples of texture features that are sufficiently successful for various tasks. In the medical domain, extensive research works have been carried out to investigate the use of the texture analysis technique in the characterisation and discrimination of biological tissue, such as liver, breast, thyroid, prostate, carotid plaques and many others, and has been proved to be valuable.

The main contributions are starting from our investigation, reported in chapter 4, was based on testing the performance of several texture-based features extracted mostly in the spatial domains of manually cropped Ultrasound ovarian tumour scan images. Training and testing the performance of some basic classifiers (SVM and kNN) with 7 such features extracted from a dataset of IOTA compliant B-mode ultrasound ovarian tumour demonstrated the effectiveness of all these machines learnt schemes in classifying Benign and malignant ovarian tumours. There was some considerable variation in the performance of these 7 schemes, but all achieved accuracy above 70% that could not be a case of random chance prediction. A close look at the nature of the predicting parameters of the various schemes helps explain the less than expected success of some schemes. In fact, within some the features and among different features, there some redundancies and correlations are easy to identify, especially within the schemes whose predicting parameters were more of statistical nature. Moreover, in these experiments, no considerations were given to variations in image quality, as a result of the speckle noise, or to variation in cropped image size.

Consulting with other experts' opinion, when a diagnostic decision cannot be made with a high degree of confidence, is a common practice by clinicians. The fusion of multiple classifiers at a different level can provide a software equivalent to this practice. Indeed, further improved accuracy was achieved by simple majority rule decision fusion of combinations of the above schemes but was not satisfactory enough to mitigate the effect of the factors mentioned above.

Chapter 5 was devoted to developing procedures that help mitigate the adverse effects of the above-mentioned performance influencing factors (i.e. Speckle noise degradation and variation in size of RoI). Removing noise are often used as a first stage enhancing the image before extracting texture features, but our investigations revealed evidence that the blanket application of de-noising might not have the desired effect in all cases. Consequently, we developed an adaptive speckle noise removal technique that adopts a smart strategy, so instead of de-noising the entire image, it acts adaptively and differently in different blocks of the same images. This strategy exploits the fact that tumour tissue contains different types of connected subregions with a concentration of different types of texture primitives, and accordingly, speckle-noise have different degradation effect in different blocks of the tumour tissues. This technique not only had the most significantly positive effect on the performance of all the above spatial domain texture-based schemes but we also demonstrated subsequently that it makes an equally significant contribution

to improve performance of texture features in non-spatial domains. The improved accuracy for all texture features by this smart approach to adaptive de-noising is supported by the experiments that are adaptively de-noising the entire tumour region which only made marginally improved accuracy to some but not all texture schemes. This approach was obviously influenced by the domain of application, but it provides more general advice that adaptive de-noising in other domains of biomedical images are likely to benefit from good knowledge of the spatial distribution of discriminating texture feature within the noisy image. In this respect, we node the important characteristic of our smart approach in that the adaptiveness criteria was obtained through training of 2 parameters.

The significant variation in tumour RoI size was the other influencing factors that were noted to have an adverse impact on the performance of some texture-based features, in particular, texture features whose parameters are of statistical nature including the LBP features. Instead of normalising the ROI sizes to one fixed size, we found that simply normalising the statistical parameters by dividing by image size helped to mitigate the adverse effect on the performance of all these types of features.

With the exception of the Gabor filter-based scheme, all the other texture features discussed above were extracted from the RoI images spatial domain. Gabor wavelet filters are frequency domain image representation, and it is by no mean is the only frequency domain texture features that have been used for image analysis. In fact, at the outset of this research project, I was aware of the work of a fellow Buckingham PhD researcher on the FFGF Fourier transform-based ultrasound ovarian tumour scan image classification, and this thesis was meant to extend that research work.

In chapter 6, I first reported our experimental work on a modified version of the FFGF scheme of Khaznadar et al (Khazendar, Shan 2016) which consist of 3 geometric parameters extracted from an elliptical shape obtained at the centre of FFT spectrum. The 2 modifications made on the original FFGA led to improved accuracy, which was mostly due to the use of our smart adaptive speckle-de-noising pre-process step when the original used another non-adaptive de-noising procedure. The obvious realisation that the 3rd geometric parameter (area) of the elliptical shape can be determined by the other 2 parameters (major and minor axes), plus other observations motivated the next experiments which showed that each of this parameter singularly perform well, and in fact the “minor axes” feature outperforms the 3-parameter FFGF scheme. These significant results highlighted an important fact about the FFT-spectrum images, in that

it has a very rich texture and structural information that are not readily discernible from its somewhat visually obscured/hidden representation. This was a strong incentive to extend the list of texture-based image feature beyond the spatial domain and beyond the FFGF schemes by extracting some of the previously defined texture features not only from the FFT-spectrum but also in any image transform domain such as the LBP domain.

The new and richer sets of texture features in the FFT-spectrum and LBP transform domains, demonstrated beyond any doubt the success of our approach and provided more evidence to support this stated hypothesis that image transformed domains provide a significantly larger pool than the spatial domain. These results are certainly one of the most interesting, perhaps surprising, and significant contributions of this thesis. The show beyond any doubt that despite their visual ambiguity, biomedical images can be analysed by ample sets of informative texture-based features that can be extracted from a variety of image representation/domains, and Machine Learning has a great role to play in early cancer detection. To some extent, this thesis has only a modest opening in the Pandora Box of image textures.

Towards the end of these investigations that were carried out in collaboration and support of clinical experts from IOTA, the time has come to put the results of these experiments into the test within a clinical setup for the eventual purpose of providing software decision support tools for gynaecologists in different hospitals and medical centres. In chapter 7, we reported on a prospective test that was designed to evaluate the performance of various versions of our ML software (but only limited to 9 texture features automatically extracted from) B-mode ovarian scan images for100 cases. The evaluation was based on comparing the simple-majority-rule predictions by various fused combinations of 7(9) spatial domain texture features with the standard gold histology-based diagnostics. The prospective test was deemed very successful when presented to the IOTA conference, and 94 % accuracy was finally achieved by version 2 with the smart adaptive pre-processing.

Although, Deep Learning tools for medical diagnostics applications started to appear on the scene during the second part of my PhD program of research, we went beyond the original objectives of the thesis and developed a pilot investigation to compare our results with the performance of the well-known ***Resnet50*** DL model trained with the original 242 IOTA dataset and tested with the 100 prospective cases. The results of this pilot testing revealed that the Deep Learning approach was outperformed by our hand-crafted 9 spatial domain features in both scenarios: the with and the without adaptive pre-

processing. One possible reason may be due to the significant variation in tumour RoI size, and DL require re-sizing the images to a fixed size. However, without much more in-depth investigations into different DL architectures, we cannot claim that no DL architecture can be found to outperform our schemes.

8.2 Future Work

Achieving optimal performance in such a complex goal requires that the scope of our reported work be further extended in several directions. Ultimately the objective should set to prepare for full-scale clinical testing to cover a much larger number of cases beyond those considered here. Ideally, future work should be dedicated to the improvement and further development of the currently proposed methods to increase overall accuracy. In particular, we will explore the following areas of research, perhaps in collaboration with other researchers.

- Firstly, extend the software used in the prospective test to be used in different hospitals for real world diagnostic environments. Am aiming to add new ideas of texture feature schemes based on FFGF, as well as a full list of features to be extracted from the FFT-spectrum as well as the LBP transformation domain. Collaborations must be expanded to test the software performance in more hospitals. Also, we should include the concept of the confidence level of decisions as part of the necessary ingredients of a computational Decision Support System (DSS) to deal with gynaecological abnormalities.

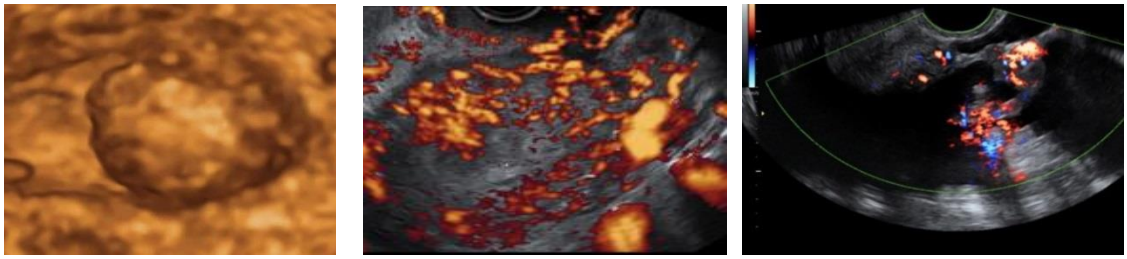
- Significant experience has been gained about the strength of many of the texture features to help develop machine learning-based trainable effective scheme for automatic segmentation of the RoI ovarian tumour images. In particular, we anticipate that we could benefit from the training process that we developed to determine an effective speckle de-noising adaptiveness criteria.

- The most important part of the future research direction is to extend our work two classes, which are Benign and Malignant ovarian tumour, into **multiclass** ovarian tumours identification. In fact, there is more interesting in distinguishing between different types of benign and malignant tumours.

- In Version 2 software used within the prospective testing, we added the two features (skewness and Kurtosis) as separate features of the statistics histogram feature that include them. This extension led to improved performance and raised the possibility of splitting other multi-parameters texture features such as the different components

of the ULBP-based features in many different ways. This specific example, of splitting features, open the way for exploiting the emerging paradigm of Topological Data Analysis which expand the analysis beyond counting the number in such LBP bins by analysing the spatial distributions in the image. A pilot study conducted, separately in collaboration with another Buckingham PhD researcher, has led to significant success in classifying benign and malignant tumours. This would be a rather extensive piece of future work that I plan to do in collaboration with Buckingham TDA research group.

- Investigate the effectiveness of using different types of US images such as colour Doppler images and 3D US scan. As shown in Figure 8-1.



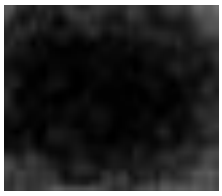
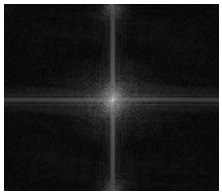
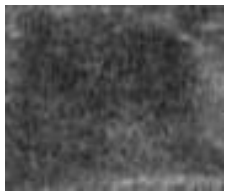
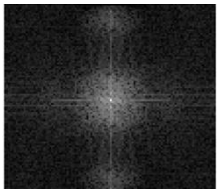
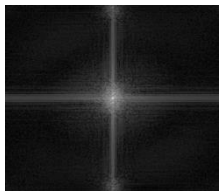
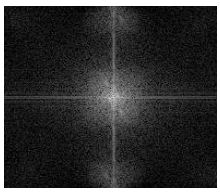
(a) (b) (c)

Figure 8- 1: Different US images (a) 3D scan of benign tumour (b) 2D colour (c) Doppler Image.

8.3 Work in Progress

I have tested the idea of FFGF (see chapter 6, section 6.3) to test the amount of solidity (where the images belong to malignant tumours contain more solids than the benign tumours) using other type of cancer (breast cancer) based on ultrasound images. The adaptive pre-processing model 2 used then applied FFT-spectrum on the de-noised image. Later, extracted one feature only (Minor axis length) for the classification purpose. Table 8-1 shows example of breast cancer for benign and malignant cases.

Table 8- 1: Samples of FFT spectrum images for breast cancer.

Benign Tumours	Its Spectrum	Malignant Tumours	Its Spectrum
			
			

The experiments were conducted by training and testing SVM classifier on our experimental dataset of 250 scan images (100 malignant and 150 benign tumours), according to the 50-50 evaluation protocol described in Chapter 4 (section4.3). Figure 8-2, below, displays the results of automatic identification of the probe set of images regarding the accuracy rate and the corresponding sensitivity and specificity measures. These results demonstrate the benefits of extracting feature vectors from the FFT domain when discriminating benign from malignant tumours.

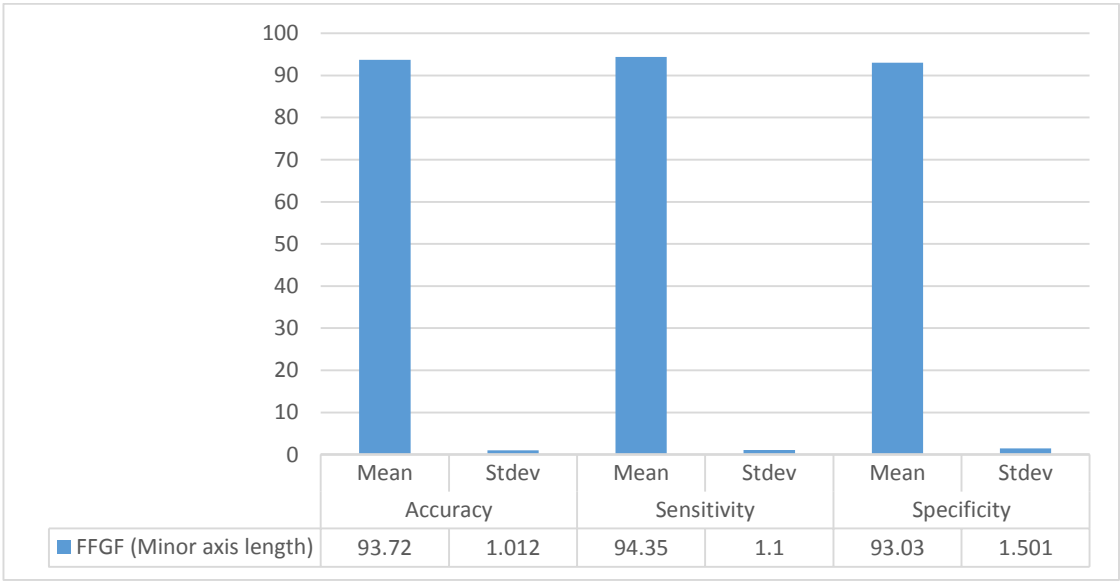


Figure 8- 2: Shows the results for the brest cancer using FFGF (Minor axis length).

The successful outcome of these investigations indicates that the FFGF based on minor axis length is given also high-performance accuracy, sensitivity and specificity when tested on another type of cancer. Future investigations will carry out using the whole FFT spectrum and other combinations of texture features.

References

Abrahim, Banazier A, Zeinab A Mustafa, Inas A Yassine, Nourhan Zayed, and Yasser M Kadah. "Hybrid total variation and wavelet thresholding speckle reduction for medical ultrasound imaging." *Journal of Medical Imaging and Health Informatics* (American Scientific Publishers) 2, no. 2 (2012): 114-124.

Acharya, U Rajendra, et al. "Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification." *Ultraschall in der Medizin-European Journal of Ultrasound* (copyright Georg Thieme Verlag KG) 35, no. 03 (2014): 237-245.

Acharya, U Rajendra, Luca Saba, Filippo Molinari, Stefano Guerriero, Jasjit S Suri, and others. "Ovarian tumor characterization and classification: a class of GyneScan™ systems." *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* San Diego: IEEE, 2012b. 4446-4449.

Al-Hassan, Nadia. *Mathematically inspired approaches to face recognition in uncontrolled conditions: super resolution and compressive sensing.* Thesis, Buckingham: University of Buckingham, 2014.

American Cancer Society. *Cancer Facts & Figures.* Atlanta: American Cancer Society, 2018.

anatomy, Human. *Female Reproductive System.* 2019. https://www.mananatomy.com/body-systems/female-reproductive-system (accessed June 2019).

Aramendia-Vidaurreta, Ver'onica, Rafael Cabeza, Arantxa Villanueva, Javier Navallas, and Juan Luis Alc'azar. "Ultrasound Image Discrimination between Benign and Malignant Adnexal Masses Based on a Neural Network Approach." *Ultrasound in medicine & biology* (Elsevier) 42, no. 3 (2016): 742-752.

Asaad, Aras and Jassim, Sabah. "Topological data analysis for image tampering detection." In *International Workshop on Digital Watermarking*, 136--146. Magdeburg: Springer, 2017.

Attlas, Nishtha and Gupta, Sheifali. "Wavelet Based Techniques for Speckle Noise Reduction in Ultrasound Images." *Nishtha Attlas et al Int. Journal of Engineering Research and Applications, ISSN*, 2014: 2248--9622.

Attlas, Nishtha, and Sheifali Gupta. "Reduction of speckle noise in ultrasound images using various filtering techniques and discrete wavelet transform: comparative

analysis." *International Journal of Research (IJR)* (Citeseer) 1, no. 6 (2014): 112-117.

Bell, Debra A. "Ovarian surface epithelial-stromal tumors." *Human pathology* (Elsevier) 22, no. 8 (1991): 750-762.

Bodytomy. September 2019. https://med.libretexts.org/Bookshelves/Anatomy_and_Physiology (accessed 9 2019).

Bovik, Alan Conrad, Marianna Clark, and Wilson S Geisler. "Multichannel texture analysis using localized spatial filters." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE) 12, no. 1 (1990): 55-73.

Burckhardt, Christoph B. "Speckle in ultrasound B-mode scans." *IEEE Transactions on Sonics and ultrasonics* (IEEE) 25, no. 1 (1978): 1-6.

Camilleri, Daniel and Prescott, Tony. "analysing the limitations of deep learning for developmental robotics." *Conference on Biomimetic and Biohybrid Systems*. Springer, 2017. 86--94.

Castellano, G, L Bonilha, LM Li, and F Cendes. "Texture analysis of medical images." *Clinical radiology* (Elsevier) 59, no. 12 (2004): 1061-1069.

Chan, Vincent, and Anahi Perlas. "Basics of ultrasound imaging." In *Atlas of ultrasound-guided procedures in interventional pain management*, 13-19. Springer, 2011.

Chaudhuri, Bidyut Baran, and Nirupam Sarkar. "Texture segmentation using fractal dimension." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE) 17, no. 1 (1995): 72-77.

Chellappa, Rama, and Shankar Chatterjee. "Classification of textures using Gaussian Markov random fields." *Acoustics, Speech and Signal Processing, IEEE Transactions on* (IEEE) 33, no. 4 (1985): 959-963.

Chen, Vivien W, et al. "Pathology and classification of ovarian tumors." *Cancer* (Wiley Online Library) 97, no. S10 (2003): 2631-2642.

Chen, Yidong, and Edward R Dougherty. "Gray-scale morphological granulometric texture classification." *Optical Engineering* (International Society for Optics and Photonics) 33, no. 8 (1994): 2713-2722.

Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* (Springer) 20, no. 3 (1995): 273-297.

Cristianini, Nello, and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge : Cambridge university press, 2000.

D.Wong, Kerry. *A Sensitive DIY Ultrasonic Range Sensor.* January 2011. http://www.kerrywong.com/2011/01/22/a-sensitive-diy-ultrasonic-range-sensor/ (accessed June 2018).

Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *International Conference on Computer Vision.* San Diego, 2005. 886-893.

Dangeti, Sarita Veera. *Denoising techniques-a comparison.* Thesis , Visakhapatnam: Louisiana State University and Agricultural and Mechanical College, 2003.

Daugman, John G. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters." *JOSA A* (Optical Society of America) 2, no. 7 (1985): 1160-1169.

Davis, Larry S. *Computer architectures for image processing.* Computer Science Department, University of Texas at Austin, 1980.

Derin, Haluk, and Howard Elliott. "Modeling and segmentation of noisy and textured images using Gibbs random fields." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE), no. 1 (1987): 39-55.

Deshpande, R., R.E. Ramalingam, N. Chockalingam, R. Naemi, H. Branthwaite, and L. Sundar. "An automated segmentation technique for the processing of foot ultrasound images." *Intelligent Sensors, Sensor Networks and Information Processing, 2013 IEEE Eighth International Conference on.* Melbourne: IEEE, 2013. 380-383.

Dhawan, Atam P. *Medical image analysis.* Vol. 31. John Wiley & Sons, 2011.

Dougherty, Geoff. *Digital image processing for medical applications.* Cambridge: Cambridge University Press, 2009.

Edler, Inge, and Kjell Lindstrom. "The history of echocardiography." *Ultrasound in medicine & biology* (Elsevier) 30, no. 12 (2004): 1565-1644.

Fishman, David A, et al. "The role of ultrasound evaluation in the detection of early-stage epithelial ovarian cancer." *American journal of obstetrics and gynecology* (Elsevier) 192, no. 4 (2005): 1214-1221.

Fix, Evelyn, and Joseph L Hodges Jr. "Discriminatory analysis-nonparametric discrimination: consistency properties." Tech. rep., DTIC Document, 1951.

G.Vasavi, and S.Jyothi. "A Comparative Study of Speckle Reduction Filters for Ultrasound Images of Polycystic Ovary." *Journal of Recent Technology and Engineering (IJRTE)* 7, no. 6S5 (2019): 2084-2089.

Gajjar, Ketan, Gemma Ogden, MI Mujahid, and Khalil Razvi. "Symptoms and risk factors of ovarian cancer: a survey in primary care." *ISRN obstetrics and gynecology* (Hindawi Publishing Corporation) 2012 (2012).

Garg, Anita, Jyoti Goal, Sandeep Malik, and Kavita Choudhary. "De-speckling of medical ultrasound images using Wiener filter and wavelet transform." *Int. J. Electron. Commun. Technol.* 2, no. 3--1 (2011).

Giger, Maryellen L, Heang-Ping Chan, and John Boone. "Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM." *Medical physics* (Wiley Online Library) 35, no. 12 (2008): 5799-5820.

Gonzalez, Rafael C, Richard E Woods, and SLU Eddins. "Digital Image Processing Using MATLAB: Pearson Prentice Hall." *Upper Saddle River, New Jersey,* 2004.

Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press. *MITPress.* 2016. http://www.deeplearningbook.org.

Goodman, Joseph W. "Some fundamental properties of speckle." *JOSA,Optical Society of America* 66 (1976): 1145--1150.

Gramellini, Dandolo, Stefania Fieni, Laura Sanapo, Giovanna Casilla, Carla Verrotti, and Giovanni B Nardelli. "Diagnostic accuracy of IOTA ultrasound morphology in the hands of less experienced sonographers." *Australian and New Zealand Journal of Obstetrics and Gynaecology* (Wiley Online Library) 48, no. 2 (2008): 195-201.

Gurjar, Preeti. *Basic of ultrasonography.* December 2017. https://biomedikal.in/2017/12/basics-of-ultrasonography/ (accessed March 2018).

Hamid, Bidi Ab. *Image texture analysis of transvaginal ultrasound in monitoring ovarian cancer.* Thesis, School of Engineering, Cardiff University, Cardiff : Cardiff University, 2011.

Hamlett, Willai C, and Thomas Koob. "Female reproductive system." *Sharks, skates, and rays: the biology of elasmobranch fishes* (The Johns Hopkins University Press Baltimore, MD), 1999: 398-443.

Han, Jiawei, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques.* Elsevier, 2011.

Hangiandreou, Nicholas J. "AAPM/RSNA Physics Tutorial for Residents: Topics in US: B-mode US: Basic Concepts and New Technology 1." *Radiographics* (Radiological Society of North America) 23, no. 4 (2003): 1019-1033.

Haralick, Robert M. "Statistical and structural approaches to texture." *Proceedings of the IEEE* (IEEE) 67, no. 5 (1979): 786-804.

Harrington, Na. *Segmentation of human ovarian follicles from ultrasound images acquired in vivo using geometric active contour models and a na{\"\i}ve Bayes classifier*. Thesis, Saskatchewan: University of Saskatchewan, 2007.

Hassner, Martin, and Jack Sklansky. "The use of Markov random fields as models of texture." *Computer Graphics and Image Processing* (Elsevier) 12, no. 4 (1980): 357-370.

Haykin, Simon and Kosko, Bart. "Gradientbased learning applied to document recognition." *IEEE*, 2009.

He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770--778. IEEE, 2016.

Hedrick, WR, and DL Hykes. "Image and signal processing in diagnostic ultrasound imaging." *Journal of diagnostic medical Sonography* (Sage Publications Sage CA: Thousand Oaks, CA) 5 (1989): 231--239.

Hijazi, Samer and Kumar, Rishi and Rowen, Chris. "Using convolutional neural networks for image recognition." *Cadence Design Systems Inc.: San Jose, CA, USA*, 2015.

Hiremath, PS, Prema T Akkasaligar, and Sharan Badiger. "Speckle Noise Reduction in Medical Ultrasound Images." *Advancements and Breakthroughs in Ultrasound Imaging. Gunarathne G.(ed.) InTech*, 2013: 201-241.

Hoskins, Peter R, Kevin Martin, and Abigail Thrush. *Diagnostic ultrasound: physics and equipment.* Cambridge University Press, 2010.

Hu, Ming-Kuei. "Visual pattern recognition by moment invariants." *IRE transactions on information theory,IEEE* 8 (1962): 179--187.

Ian Goodfellow, Bengio, Yoshua and Courville, Aaron. *Deep learning. Book in preparation for MIT Press.* http://www. deeplearningbook. org, 2016.

Ibrahim, Dheyaa Ahmed. *Multi-level Segmentation of Gynaecological Ultra- sound Images using Texture-based Trainable Models.* Thesis, Buckingham: University of Buckingham, 2018.

Ibrahim, Dheyaa Ahmed, et al. "Automatic segmentation and measurements of gestational sac using static B-mode ultrasound images." *Mobile Multimedia/Image Processing, Security, and Applications.* Baltimore: SPIE, 2016. 98690B.

Ibrahim, Dheyaa Ahmed, Hisham Al-Assam, Sabah Jassim, and Hongbo Du. "Multi-level Trainable Segmentation for Measuring Gestational and Yolk Sacs from Ultrasound Images." *Annual Conference on Medical Image Understanding and Analysis*, no. Springer (2017): 86--97.

Ilonen, Jarmo and Kamarainen, Joni-Kristian, Pekka Paalanen, Miroslav Hamouz, Josef Kittler, and Heikki Kalviainen. "Image feature localization by multiple hypothesis testing of Gabor features." *IEEE Transactions on Image Processing* 17 (2008): 311--325.

Ilonen, Jarmo, Joni-Kristian Kämäräinen, and Heikki Kälviäinen. "Efficient computation of Gabor features." *Lappeenranta University of Technology* (Elsevier) 29, no. 12 (2005).

Julesz, Bela. "Experiments in the visual perception of texture." *Scientific American* 232 (1975): 34-43.

Kaijser, Jeroen. "Towards an evidence-based approach for diagnosis and management of adnexal masses: findings of the International Ovarian Tumour Analysis (IOTA) studies." *Facts, views \& vision in ObGyn* (Vlaamse Vereniging voor Obstetrie en Gynaecologie) 7, no. 1 (2015): 42.

Kaplan, Lance M, and CC Jay Kuo. "Texture roughness analysis and synthesis via extended self-similar (ESS) model." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE) 17, no. 11 (1995): 1043-1056.

Khazendar, S, et al. "Automated classification of static ultrasound images of ovarian tumours based on decision level fusion." *Computer Science and Electronic Engineering Conference (CEEC), 2014 6th.* Colchester: IEEE, 2014. 148-153.

Khazendar, Shan. *Computer-Aided Diagnosis of Gynecological Abnormality using B-mode Ultrasound Images.* Thesis, Buckingham: University of Buckingham, 2016.

Khazendar, Shan, et al. "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* (Vlaamse Vereniging voor Obstetrie en Gynaecologie) 7, no. 1 (2015): 7.

Khiyal, Malik Sikandar Hayat, Aihab Khan, and Amna Bibi. "Modified Watershed Algorithm for segmentation of 2D images." *Issues in Informing Science and Information Technology* 6 (2009): 877-886.

Kinkel, Karen, Hedvig Hricak, Ying Lu, Kyo Tsuda, and Roy A Filly. "US characterization of ovarian masses: a meta-analysis." *Radiology* (Radiological Society of North America) 217 (2000): 803-811.

Kourou, Konstantina and Exarchos, Themis P and Exarchos, Konstantinos P and Karamouzis, Michalis V and Fotiadis, Dimitrios I. "Machine learning applications in cancer prognosis and prediction." *Computational and structural biotechnology journal* (Elsevier) 13 (2015): 8--17.

Krishnamachari, Santhana, and Rama Chellappa. "Multiresolution Gauss-Markov random field models for texture segmentation." *Image Processing, IEEE Transactions on* (IEEE) 6, no. 2 (1997): 251-267.

Kullback, Solomon. *Information theory and statistics.* Courier Corporation, 1997.

Kumar, Girish, and Pradeep Kumar Bhatia. "A detailed review of feature extraction in image processing systems." *Fourth international conference on advanced computing & communication technologies.* Rohtak: IEEE, 2014. 5-12.

Kupesic, S, and BM Plavsic. "Early ovarian cancer: 3-D power Doppler." *Abdominal imaging* (Springer) 31 (2006): 613-619.

Kurani, Arati S, Dong-hui Xu, Jacob Furst, and Daniela Stan Raicu. "Raicu. Co-occurrence Matrices for Volumetric Data." *7th IASTED International Conference on Computer Graphics and Imaging, Kauai.* USA: Citeseer, 2004.

Kurtz, Alfred B, et al. "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* (Radiological Society of North America) 212, no. 1 (1999): 19-27.

Landgren, Matilda, Niels Christian Overgaard, and Anders Heyden. "Segmentation of the left heart ventricle in ultrasound images using a region based snake." *SPIE Medical Imaging.* Florida: SPIE, 2013. 866945-866945.

Lei, Yi-ming, Xi-mei Zhao, and Wei-dong Guo. "Cirrhosis recognition of liver ultrasound images based on SVM and uniform LBP feature." *2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).* Chongqing: IEEE, 2015. 382-387.

Lemaitre, Guillaume, Robert Marti, Jordi Freixenet, Joan C Vilanova, Paul M Walker, and Fabrice Meriaudeau. "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review." *Computers in biology and medicine* (Elsevier) 60 (2015): 8-31.

Lerski, RA, K Straughan, LR Schad, D Boyce, S Bl{\"u}ml, and I Zuna. "VIII. MR image texture analysis—an approach to tissue characterization." *Magnetic resonance imaging* (Elsevier) 11, no. 6 (1993): 873-887.

Levine, Martin D. *Vision in man and machine.* Vol. 574. McGraw-Hill New York, 1985.

Li, Chun Hung and Lee, CK. "Minimum cross entropy thresholding." *Pattern recognition* (Elsevier) 26 (1993): 617--625.

Lu, Chun S, Pau C Chung, and Chih F Chen. "Unsupervised texture segmentation via wavelet transform." *Pattern Recognition* (Elsevier) 30, no. 5 (1997): 729-742.

Mallat, Stephane G. "Multifrequency channel decompositions of images and wavelet models." *Acoustics, Speech and Signal Processing, IEEE Transactions on* (IEEE) 37, no. 12 (1989): 2091-2110.

Mandelbrot, Benoit B. "The fractal geometry of nature/Revised and enlarged edition." *New York, WH Freeman and Co., 1983, 495 p.*, 1983.

Manjunath, BS, and Rama Chellappa. "Unsupervised texture segmentation using Markov random field models." *IEEE Transactions on Pattern Analysis \& Machine Intelligence* (IEEE), no. 5 (1991): 478-482.

Materka, Andrzej, Michal Strzelecki, and others. "Texture analysis methods--a review." *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, 1998: 9-11.

Mathias, JM, PS Tofts, and NA Losseff. "Texture analysis of spinal cord pathology in multiple sclerosis." *Magnetic Resonance in Medicine* (Wiley Online Library) 42, no. 5 (1999): 929-935.

Mayerhoefer, Marius E, Martin Breitenseher, Gabriele Amann, and Martin Dominkus. "Are signal intensity and homogeneity useful parameters for distinguishing between benign and malignant soft tissue masses on MR images?: Objective evaluation by means of texture analysis." *Magnetic resonance imaging* (Elsevier) 26, no. 9 (2008): 1316-1322.

Media, Alila Medical. *Cancer cells in a growing tumor.* 2019. https://www.shutterstock.com/image-illustration/cancer-cells-growing-tumor-107362892 (accessed May 2019).

Medical Image Modalities . Spetumber 2019.

Metha, Sandip. "Speckle Noise Reduction Using Hybrid Wavelet Packets-Wiener Filter." *International Journal of Computer Sciences and Engineering* 5, no. 10 (2017): 95-99.

Modos, DezsHo, et al. "Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies." *NPJ systems biology and application,Nature Publishing Group* 3 (2017): 2.

Morris, DT. "An evaluation of the use of texture measurements for the tissue characterisation of ultrasonic images of in vivo human placentae." *Ultrasound in Medicine and Biology* (Elsevier) 14, no. 5 (1988): 387-395.

Nailon, William Henry. "Biomedical Imaging." Chap. 4 in *Texture analysis methods for medical image characterisation*. China: INTECH Open Access Publisher, 2010.

Nobi, MN and Yousuf, MA. "A New Method to Remove Noise in Magnetic Resonance and Ultrasound Images." *Journal of scientific research* 3 (2011).

Ojala, T., M. Pietikäinen and D. Harwood. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition, Elsevier*, 1996: 51-59.

Ojala, Timo, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition* (Elsevier) 29, no. 1 (1996): 51-59.

oplin, Ryan and Varadarajan, Avinash V and Blumer, Katy and Liu, Yun and McConnell, Michael V and Corrado, Greg S and Peng, Lily and Webster, Dale R. "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning." *Nature Publishing Group* 2 (2018): 158.

Ozgen, Canan. *A Medical Image Processing and Analysis Framework.* Ph.D. dissertation, MIDDLE EAST TECHNICAL UNIVERSITY, middle east technical university, 2011.

Papoulis, Athanasios, and S Unnikrishna Pillai. *Probability, random variables, and stochastic processes.* Tata McGraw-Hill Education, 2002.

Pentland, Alex P. "Fractal-based description of natural scenes." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE), no. 6 (1984): 661-674.

Permuth-Wey, Jennifer, and Thomas A Sellers. "Epidemiology of ovarian cancer." In *Cancer epidemiology*, 413-437. Springer, 2009.

Pietikainen, Matti, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. "Local binary patterns for still images." In *Computer vision using local binary patterns*, 13-47. Springer, 2011.

Powers, Jeff, and Frederick Kremkau. "Medical ultrasound systems." *Interface focus* (Royal Society) 1, no. 4 (2011): 477-489.

Quaia, Emilio. *Contrast media in ultrasonography.* Springer, 2005.

Rafiee, Ali, Ahad Salimi, and Ali Reza Roosta. "A novel prostate segmentation algorithm in TRUS images." *World Academy of Science, Engineering and Technology* 45 (2008): 120-124.

Rosenfeld, Azriel, and Joan S Weszka. *Picture recognition.* Springer, 1976.

Russakovsky, Olga and Deng, Jia and Su, Hao and Krause, Jonathan and Satheesh, Sanjeev and Ma, Sean and Huang, Zhiheng and Karpathy, Andrej and Khosla, Aditya and Bernstein, Michael and others. "Imagenet large scale visual recognition challenge." *International journal of computer vision, Springer* 115 (2015): 211--252.

Saranya, R and Maheswari, S Uma. "Follicle detection in ovary image using adaptive particle swarm optimization." *Journal of Medical Imaging and Health Informatics* 6, no. American Scientific Publishers (2016): 125--132.

Scully, Robert, and Leslie H Sobin. *Histological typing of ovarian tumours.* Vol. 9. New York: Springer Science & Business Media, 1999.

Serra, Jean. *Image analysis and mathematical morphology, v. 1.* FL: Academic press, 1983.

Sivakumar, R and Nedumaran, D. "Comparative study of speckle noise reduction of ultrasound b-scan images in matrix laboratory environment}." *International Journal of Computer Applications* 10 (2010): 46--50.

Smyth, PP, JE Adams, RW Whitehouse, and CJ Taylor. "Application of computer texture analysis to the Singh Index." *The British Journal of Radiology* (The British Institute of Radiology) 70, no. 831 (1997): 242-247.

Sohail, Abu Sayeed Md, Md Mahmudur Rahman, Prabir Bhattacharya, Srinivasan Krishnamurthy, and Sudhir P Mudur. "Retrieval and classification of ultrasound images of ovarian cysts combining texture features and histogram moments." *IEEE International Symposium on Biomedical Imaging: From Nano to Macro.* Rotterdam: IEEE, 2010. 288-291.

Springenberg, Jost Tobias and Dosovitskiy, Alexey and Brox, Thomas and Riedmiller, Martin. "Striving for simplicity: The all convolutional net." *arXiv preprint arXiv:1412.6806*, 2014.

Szczypinski, Piotr M, Michal Strzelecki, Andrzej Materka, and Artur Klepaczko. "MaZda—a software package for image texture analysis." *Computer methods and programs in biomedicine* (Elsevier) 94, no. 1 (2009): 66-76.

Tamilkudimagal, D, and ME Kalpana. "Squeeze box filter for contrast enhancement in ultrasound despeckling." *International Conference on Emerging Trends in Electrical and Computer Technology.* Nagercoil, India: IEEE, 2011. 524-530.

Tan, David SP, Roshan Agarwal, and Stanley B Kaye. "Mechanisms of transcoelomic metastasis in ovarian cancer." *The lancet oncology* (Elsevier) 7, no. 11 (2006): 925-934.

Tangir, JACOB, and PETER E Schwartz. "Fertility preservation in the management of germ cell ovarian cancer." *CME JOURNAL OF GYNECOLOGIC ONCOLOGY* (PRIMED-X-PRESS) 8, no. 2 (2003): 117-120.

Thijssen, JM and Oosterveld, BJ and Hartman, PC and Rosenbusch, GJE. "Correlations between acoustic and texture parameters from RF and B-mode liver echograms." *Ultrasound in Medicine and Biology, Elsevier* 19, no. 1 (1993): 13--20.

Timmerman, Dirk, et al. "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* (Wiley Online Library) 36, no. 2 (2010): 226-234.

Timmerman, Dirk, et al. "Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group." *BMJ* (BMJ Publishing Group Ltd) 341 (2010).

Tingulstad, Solveig, et al. "Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses." *BJOG: An International Journal of Obstetrics \& Gynaecology* (Wiley Online Library) 103, no. 8 (1996): 826-831.

Togashi, Kaori. "Ovarian cancer: the clinical role of US, CT, and MRI." *European radiology* (Springer) 13, no. 6 (2003): L87--L104.

Twickler, Diane M, and Elysia Moschos. "Ultrasound and assessment of ovarian cancer risk." *American Journal of Roentgenology* (Am Roentgen Ray Soc) 194, no. 2 (2010): 322-329.

UK Cancer Research, U. *Types of ovarian cancer.* December 2018. https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/types (accessed July 2017).

Unger, Katharina. *Evaluation of Machine Learning Frameworks on Tuberculosis Classification of Chest Radiographs.* Bachelor of Science, TU Wien: Faculty of Informatics, 2017.

Valentin, Lil, et al. "Ultrasound characteristics of different types of adnexal malignancies." *Gynecologic oncology* (Elsevier) 102 (2006): 41-48.

Valkealahti, Kimmo, and Erkki Oja. "Reduced multidimensional co-occurrence histograms in texture classification." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (IEEE) 20, no. 1 (1998): 90-94.

Van Calster, Ben, et al. "Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors." *Facts, views & vision in ObGyn* (Vlaamse Vereniging voor Obstetrie en Gynaecologie) 7, no. 1 (2015): 32.

Veit, Andreas and Wilber, Michael J and Belongie, Serge. "Residual networks behave like ensembles of relatively shallow networks." In *Advances in neural information processing systems*, 550--558. Cornell University, 2016.

Wang, Yi, et al. "Studies on tissue characterization by texture analysis with co-occurrence matrix method using ultrasonography and CT imaging." *Journal of Medical Ultrasonics* (Springer) 29, no. 4 (2002): 211-223.

Webb, Andrew R. *Statistical pattern recognition.* Sussex: John Wiley & Sons, 2003.

Weldon, Thomas P, William E Higgins, and Dennis F Dunn. "Efficient Gabor filter design for texture segmentation." *Pattern Recognition* (Elsevier) 29, no. 12 (1996): 2005-2015.

Wiener, Norbert. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications.* MIT Press, 1949.

Xian, Guang-ming. "An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM." *Expert Systems with Applications* (Elsevier) 37, no. 10 (2010): 6737-6741.

Xie, Jun, Yifeng Jiang, H-T Tsui, and P-A Heng. "Boundary enhancement and speckle reduction for ultrasound images via salient structure extraction." *IEEE transactions on biomedical engineering* (IEEE) 53, no. 11 (2006): 2300-2309.

Yen, Philip, Kathleen Khong, Ramit Lamba, Michael T Corwin, and Eugenio O Gerscovich. "Ovarian Fibromas and Fibrothecomas Sonographic Correlation With Computed Tomography and Magnetic Resonance Imaging: A 5-Year Single-Institution Experience." *Journal of Ultrasound in Medicine* (Am inst Ulrrasound Med) 32, no. 1 (2013): 13-18.

Zhang, Yu-Jin. *Advances in image and video segmentation.* Beijing: IGI Global, 2006.

Zhu, Changming and Ni, Jun and Li, Yanbo and Gu, Guochang. "Speckle noise suppression techniques for ultrasound images." *Fourth International Conference on Internet Computing for Science and Engineering* (IEEE), 2009: 122--125.