# Multi-level Segmentation of Gynaecological Ultrasound Images using Texture-based Trainable Models

By

Dheyaa Ahmed Ibrahim

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

November, 2018

# Abstract

Ultrasound imagery has been widely used for different medical diagnoses. The purpose of such technology is to provide a picture of internal body tissue structures to monitor, diagnose, or treat medical conditions. Unlike X-rays and other radiographic scanning systems, ultrasound scanning is non-intrusive and has no side effects. Hence it is repeatedly used throughout pregnancy to identify potential abnormalities associated with the first trimesters of pregnancy, especially the risk of miscarriage. In addition, it is also one of the most crucial considerations during the assessment of a variety of ovarian tumours.

It is critical to acknowledge that the quality of a medical diagnosis based on manual measurements can be influenced by the inconsistency of the measurements obtained by different gynaecologists, and indeed even by the same gynaecologist (i.e. inter- and intra-observer variabilities). Advances in image processing technology and the emergence of machine learning as a tool for image analysis have significantly increased the potential of developing automated tools for the enhancement of diagnostic accuracy in general.

The overall aim of this thesis is to develop and test the performance of novel automated machine learning solutions that analyse B-mode ultrasound images of the Gestational Sac (GS) and the ovary to detect and classify gynaecological anomalies that have health implications for women. This thesis particularly addresses the segmentation problem of the GS during pregnancy, which is an essential step of automatic diagnostic systems. Accordingly, the research investigations reported in this thesis has been primarily concerned with the development of novel and automatic tools for segmenting a given region of interest, before turning our attention to the feature extraction step to help determine the most discriminating parameters for the identification of abnormality during pregnancy. This work has great deal of synergy with ovarian tumour diagnostic systems, and we shall accordingly illustrate these synergies.

We have evaluated various traditional segmentation techniques to highlight the factors that influence the associated performance and the shortcomings. Based on the evaluation, we proposed a hybrid solution that combines wavelet-based enhancement and a simple threshold with object detection model to locate the region of interest (GS), which is less complicated problem than locating ovarian tumour masses. Identifying the causes of the limitations of the threshold-based Ultrasound ovarian scan segmentation, provided convincing arguments to the investigation innovative machine learning based trainable segmentation solutions.

The machine learning approach (Neural Network (NN) and Support Vector Machine (SVM)) for segmentation and object detection has been used in different applications, i.e. face detection. This thesis demonstrates the suitability of this approach for ultrasound image segmentation. We designed an effective multi-level segmentation of the Gestational and Yolk Sacs using texture-based trainable models for pregnancy assessment from ultrasound images. To avoid the over and under-segmentation problems resulting from trainable segmentation, we further tweaked this approach in two ways:1) a trainable Region growing scheme, and 2) an object detection model based on the Cascade model. The thesis also demonstrates that combining this approach with the watershed transform helps successfully segmentation of unilocular and multilocular cysts from ultrasound images of ovarian tumours, which is essential for identifying tumour type. Finally, and to demonstrate the effectiveness of the proposed segmentation techniques, different texture-based features were extraction to identify the abnormality for both a pregnancy and ovarian tumour. The thesis argues that machine learning diagnosis using texture features has great potential to capture the signs of abnormalities in US images for both pregnancy and ovarian tumour. Experimental results on ovary ultrasound images (for both Miscarriage and Ovarian cases) demonstrate the effectiveness of the proposal solutions in producing accurate measurements as well as identify the abnormal cases in early stage.

# Acknowledgment

During my research, I received advice and support from many people, who I would like to acknowledge here.

I would like to thank my supervisors Prof. Sabah Jassim, Mr Hongbo Du and Dr. Hisham Al-Assam for their support and guidance.

Special thanks to Dr. Jessica Farren, Prof. Tom Bourne and Dr. Ahmed Sayasneh for providing the data set and guiding with medical information.

I should not forget to acknowledge all my lab members and colleagues for their advice and their willingness to share their bright thoughts with me, for shaping up my research.

Lastly, the financial support from the government of Iraq / The Higher Committee for Education Development in Iraq (HCED), who granted my study, is gratefully acknowledgement.

**Dheyaa Ahmed Ibrahim**

**Dedication**

This work is humbly dedicated to all my valuable treasures in my life:

To my parents, brothers and sisters,

To my beautiful wife, Esraa Khalid

**Dheyaa Ahmed Ibrahim**

# Abbreviations

| | |
|---|---|
| ANN | Artificial Neural Network |
| B-mode | Brightness-mode |
| CA-125 | Cancer Antigen 125 |
| CAD | Computer Aided Diagnosis |
| CRL | Crown-Rump Length |
| FD | Fractal Dimension |
| FN | False Negative |
| FP | False Positive |
| GLCM | Grey-Level Co-Occurrence Matrix |
| GS | Gestational Sac |
| HOG | Histogram of Orientation Gradient |
| IOTA | International Ovarian Tumor Analysis |
| LBP | Local Binary pattern |
| MSD | Mean Sac Diameter |
| PUV | Pregnancy Unknown Viability |
| RG | Region Growing |
| ROC | Receiver Operating Characteristic |
| ROI | Region of Interest |
| SVM | Support Vector Machine |
| TN | True Negative |
| TP | True Positive |
| ULBP | Uniform Local Binary Pattern |
| US | Ultrasound |
| YS | Yolk Sac |

**Table of Contents**

**List of Figures**

**Tables**

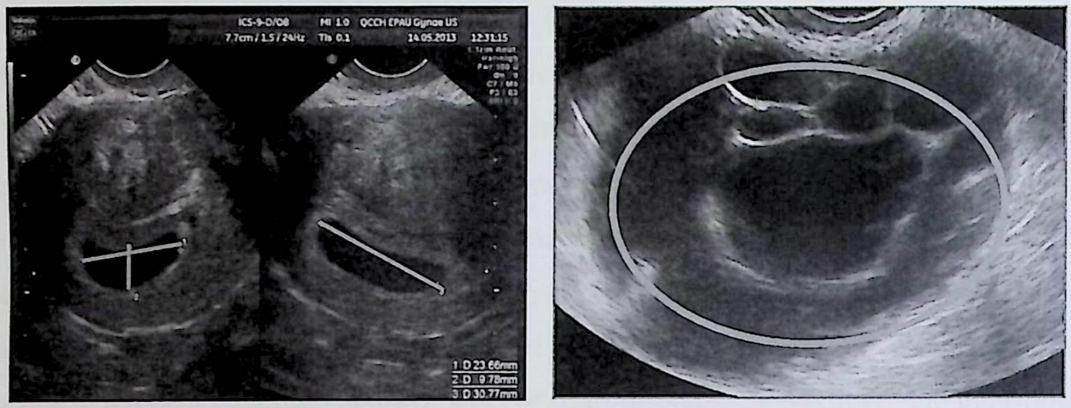
# Chapter 1. **Introduction**

Medical imaging techniques include a variety of modalities such as ultrasound (US), Magnetic Resonance (MR), conventional X-Ray, computer tomography (CT scan), etc. These techniques are based on result of scanning tissues, organs and bones using, respectively, ultrasound waves, X-ray radiation, and magnetic and radio waves that passes through the body parts for medical examination. The outcome are images that clinicians and radiologist can use to assess and diagnose abnormalities, if any, and decide the nature of treatments.

Ultrasound imaging is arguably the most commonly deployed medical image modality, and has been used for over half a century. Due to its safe and non-intrusive nature, this imaging modality is frequently used in the field of gynaecology, particularly during the period of pregnancy for checking the baby's growth (Michailovich and Tannenbaum 2006). Much of the existing research has shown that ultrasound imaging has little or no obvious negative effects on child hearing, cancers and birth weight (Torloni, et al. 2009). With ultrasound imaging technology, the evaluation of gestation within the first three months of pregnancy usually help confirm the presence and the number of the pregnancies, together with the location and well-being of the embryo. Ultrasound scan of the ovary is also used as a powerful non-intrusive tool for examining suspicious masses for gynaecological abnormalities including early diagnosis of ovarian cancer. This thesis is generally concerned with the analysis of ultrasound images for both problem domains aiming to exploit advanced machine learning technologies to provide innovative solutions for the growing challenges to the healthcare systems. Our reported research activities are more focussed on the analysis of US scan images of the gestational sac for diagnosing miscarriages, but will also report on implications of those investigations for the analysis of US scan images of ovarian tumours for signs of malignancy.

## 1.1 The Specific Research Problem Under Investigation

Pregnant women are routinely examined in the first trimester of pregnancy and an ultrasound scan of the ovary is used as an initial assessment of the state of pregnancy. This scan is meant to examine the Gestational Sac (GS), in two scanning planes (sagittal and transverse planes). The first element of test is to measure the size of the GS as an indicator of its state of health. The Mean Sac Diameter (MSD) of the GS is generated from three diameters measured in the sagittal and transverse planes (See Figure 1-1-A). MSD is a frequently used measure of the GS size, indicating if the early pregnancy is normal or at risk of miscarriage (empty large GS with MSD ≥ 25mm. See Chapter 2 for more details). Currently, in clinics and hospitals, GS measurements are done manually by the ultrasound scan operator.



A) MSD measurements for the pregnancy case B) Ovarian tumour mass

***Figure 1-1: Shows the manual measurement for both pregnancy and ovarian tumour: A) shows the MSD measurement for sagittal and transverse planes and B) ovarian tumour mass includes multilocular cyst***

Another closely related topic of research in the gynaecology field is that of the accurate and early diagnosis of ovarian cancer. According to the first prospective study in (Braem, et al. 2012), multiple miscarriages are also associated with an increased risk of ovarian cancer. Nevertheless, detecting ovarian cancer in early and treatable stages still remains an arduous task (NHS 2015), and hence has attracted a lot of attention from researchers in different disciplines. A B-mode ultrasound imaging can depict morphological features including unilocular or multilocular cysts, presence of fluid, solid tissues, internal wall structures, papillary projections and acoustic shadows. A superimposed Doppler image on top of a B-mode image reveals blood flow information within tumour areas, offering additional assistance in tumour diagnosis. The results of combining these different types of information enable clinicians to determine the seriousness of the tumour (Sayasneh, et al. 2015) (Sayasneh, et al. 2016). Figure 1-1-B shows ovarian tumour mass includes multilocular cyst.

Therefore, we are interested in common procedures needed for the analysis of ultrasound images for both problem domains. Although the primary focus of our research is the analysis of US scan images of the gestational sac for diagnosing miscarriages, the outcome of such research will also be tested for its impact on the analysis of US scan images of ovarian tumours for signs of malignancy.

The most common task in both areas of research is the segmentation of the Region of Interest (ROI). In the case of pregnancy, this relates to the automatic detection of the GS, whereas in the case of other gynaecological abnormality diagnosis, including ovarian tumour, it is concerned with detecting the suspicious masses.

Manual segmentation of the ROI in these cases is conducted by well-trained gynaecologist and radiologists. Manual segmentation may involve multiple subjective measurement decisions that will increase the possibility of inter- and intra-observer errors. Errors in judging miscarriage cases in early pregnancy and ovarian tumour can have severe consequences in terms of both missed opportunities (false negatives) and false alarms (false positives). In particular, some clinical practitioners have argued strongly for intolerance towards 'false alarms' (Bourne 2016) (Abdallah, Daemen and Guha, et al. 2011). Therefore, developing automatic solutions add value in terms of speeding up the process of analysis and reducing intra- and inter-observer variation problems.

Recent advances in machine learning is an incentive to automate this process for the sac of enhancing efficiency of US scan image analysis and reducing the burden on healthcare systems. Automatic medical image analysis, in general, is hindered by a variety of factors relating to the nature of imaging modality as well as the variation of the quality of scan images which in turn present difficult technical challenges (McInerney, et al. 2002). For ultrasound images, although various techniques have been developed for segmenting ROI and a lot of progress has been made (Meiburger, Acharya and Molinari 2017), the existing solutions are still hindered by severe technical difficulties (at least until now). These challenges can be summarised as follows:

- The shapes of soft tissues inside the human body are not only complex but also highly variable due to different displacements of the human body when the image is acquired.
- The ROI may be of an irregular shape. The irregularity of the ROI makes the filtering of non-ROI objects based on geometry features extremely difficult.
- Ultrasound images tend to be low in contrast, with blurry boundaries to the different objects present (Noble and Boukerroui 2006). The pixel intensity values in the boundary region are quite similar, which makes it difficult to identify the precise border between the ROI and the background (Over-segmentation problem), even for trained human operator. Traditional boundary-finding algorithms based on gradient information will fail in such cases.
- Inhomogeneity within the ROI means that it may contain areas of different textures within its boundary. Such inhomogeneity within the ROI itself may in some confusion

between the actual ROI and its background areas outside the ROI, causing a problem known as "under-segmentation", i.e., parts of the ROI are considered to be the background and are thus erroneously excluded from the final segmentation results.

- Like many other types of medical images, ultrasound images are inevitably noisy. Certain noise types are multiplicative in nature, making them hard to remove using conventional filters. On the other hand, it is highly likely that a selected filter not only removes the noise but also crucial details of the image that may be distinct for the object within the image.
- Shadow is a frequent phenomenon occurring in ultrasound images. In the problem domains of our research interest, shadows frequently occur in the images of benign tumours. Unfortunately, shadows are normally of irregular shape, and they usually lay across the ROI and its background, making the task of segmentation even more complicated.

Traditional approaches to automatic segmentation i.e. threshold, watershed transform, region growing, and edge detection-based techniques are adversely affected by a number of factors including quality of the ultrasound image, poor and blurred ROI border. Based on the evaluations and investigations, we found that those factors have made the segmentation task difficult due to three problems1) missing the correct ROI, 2) over- segmentation and 3) under-segmentation. Therefore, our investigations explore the alternative ways of using trainable approaches in machine learning to determine gynaecological anomalies, thus supplementing domain expert diagnostic decisions.

The next step in automatic imaged-based diagnostics relates to feature extraction and the use of appropriate classification tools to determine the nature of the relevant abnormality. Two type of features have been extracted in this step, first, the morphological features (automated the existing manual measurement i.e. MSD measurements). Second, image texture features of the segmented ROI such as the LBP features. The texture of ultrasound images has been carried out in many types of research to characterise the echo-texture of B-mode images quantitatively. The principles of analysing the texture of B-mode ultrasound imaging are, if disease procedures affect the structure of the tissue, the tissue will reflect a changed in ultrasound signal, which will give different texture features value to the normal tissue (Morris 1988). Based on that, it is expected that texture features derived from abnormal and normal tissues will be different. Therefore, this work will be focused on two points to identify the miscarriage cases in early stages based on the texture information. First, extracted the texture features from ultrasound images of GS offer any values for diagnosis of miscarriage cases besides MSD. Second, can such feature vectors help in confirming MC cases (MSD≥25mm), i.e. offering any added value. Two references points of diagnosis ground truth have been used: (a) Diagnosis on presented scan, and (b) ultimate diagnose.

Finally, those features are fed into the classifier to identify the abnormality of the miscarriage as well as the ovarian tumour type. The success or failure of automatic segmentation could have significant impact on reliability of automatic decisions. The choice of the appropriate classifiers is influenced by many factors including the dimensionality of the features, the size of available samples, as well as robustness against variations in feature/image quality. Generally, image/feature quality is dealt with by image enhancement and de-noising procedures. In general, Ultrasound images suffer from speckle noise and most researchers use blanket speckle de-noising, but we shall test the usefulness of such an approach and develop an adaptive alternative. We shall use different classifiers for different tasks.

## 1.2 Research Aim and Objectives

The overall aim of this thesis is to develop and test the performance of novel automated machine learning solutions that analyse B-mode ultrasound images of the GS and the ovary to detect and classify gynaecological anomalies that have health implications for women. In the context of a Computer-aided diagnosis (CAD) system that explained in detail in chapter 2 (see section 2-1), the research specifically focusses on a particularly crucial and difficult step in this system: segmentation of ROI. The relevant research questions are (a) whether it is feasible to automatically segment the ROI for accurate measurements of morphological features from the input images, and what the limitations with the existing methods are, and (b) whether the segmented ROIs are effective for detecting miscarriage cases and for identifying benign and malignant ovarian tumours. To address these research questions, the following research objectives have been set:

1. Investigating the problems and difficulties in segmenting the GS and Yolk Sac (YS) from B-mode ultrasound images, and developing automated methods for effectively locating and measuring the GS and YS. Successful outcomes of this aspect of research will lead to the development of a fully automated solution for measuring GS and YS sizes and consequently making correct predictions about the state of pregnancy (miscarriage case (MC) versus pregnancy of unknown viability (PUV)). The technical challenges faced by this aspect of the research is the problem of over- and under-segmentation of ROIs (see section

1-1). We are particularly interested in trainable approaches for segmentation rather than the more conventional threshold-based solutions.

2. Investigating the suitability of the developed and tested solutions for Objective 1 in another problem domain, i.e., segmentation of ovarian tumour masses from B-mode ultrasound images, and developing more advanced solutions in overcoming limitations of the previous solutions for extracting tumour masses. Ovarian tumour structures are known as being more complex due to the similarity between the textures of the ROI and those for the background area of the image. This complexity leads to difficulties in isolating the correct tumour areas. We are again particularly interested in using trainable approaches to solving difficult cases and combining these trainable approaches with more conventional region-growth methods. Despite the anticipated difficulties, successful outcomes of this aspect of research can be used to evaluate the extent to which the underlying principles can be of assistance in solving another complex problem in a completely different domain of application, and to make significant progress towards automated solutions for ovarian mass diagnosis.
3. Investigating the effects of speckle noise in B-mode ultrasound images towards accurate segmentation of ROI and correct classification of gynaecological anomalies, and developing effective solutions to suppress this type of noise so as to improve segmentation and diagnosis accuracy. The effective isolation of the speckle noise and reduction of its levels using appropriate filtering techniques and approaches is important in enhancing the input ultrasound images without losing key features. We are particularly interested in adaptive filtering solutions that can determine how much noise suppression should be applied to different local regions of the image. Successful outcomes may further improve the outcomes of the previous two objectives, providing a better understanding of the effect of denoising speckle noise.
4. Evaluating the effectiveness, and the extent of the effectiveness, of all proposed solutions within a CAD system context by (a) identifying miscarriage cases during the early stages of pregnancy based on both morphological features and texture features, and (b) classifying benign and malignant ovarian tumours. In other words, the effectiveness of the proposed solutions can only become meaningful in the CAD system context of decision making. In the evaluation work (a), we are particularly interested in knowing if the texture features extracted from the successfully segmented GS can offer any additional diagnostic value to identifying cases of miscarriage in addition to known morphological features such as MSD.

It is necessary to emphasise the importance of the segmentation step in a CAD system. Effective segmentation of ROI is crucial for the automation and diagnostic accuracy of the CAD system. Successful segmentation should facilitate the complete automation of the CAD system (Shan 2011), which in turn minimizes the effects of the operator-dependent nature inherent to ultrasound imaging (Hwang, et al. 2005), and further makes the diagnosis process reproducible. The accuracy of segmentation is essential to extracting crucial geometry and texture features within the boundary of the segmented ROI to discriminate abnormalities from normal cases (Shan 2011).

## 1.3 Research Methodology

As described in Section 1.2, the main aim of this thesis is the automated diagnosis of gynaecological anomalies. Although the current research focusses on the issue of ROI segmentation, it will also be concerned with various aspects of computer vision and machine learning. Any proposed ROI segmentation methods will be tested and evaluated against the CAD system framework. In other words, the research is not only about how closely the segmented ROIs match the actual objects of interest but is also about the extent to which ROI segmentation results affect the outcome of the final diagnosis. It is for this reason that we need to reiterate our research focusses throughout a CAD workflow, from start to end.

### 1.3.1 Positioning the Research within a CAD System Workflow

Figure 1-2 positions this research within the context of a CAD system. This research on the segmentation of ROI serves two purposes: first, the segmented ROI is automatically determined to obtain morphological features such as the MSD of the GS size during early pregnancy, and further whether a given cyst is unilocular or multilocular; these features may then be used for diagnosis. This form of computer-aided diagnosis is called CAD1 in Figure 1-2. Second, the segmented ROI is passed to the next step of a fully automated CAD system, i.e., feature extraction, to obtain image-based texture features for classification. This route is called CAD2, as shown in Figure 1-2.

Usually, ultrasound images are corrupted by random noise during data acquisition. Such noise can make the segmentation task extremely difficult. For this reason, image processing filters should be applied to remove such noise and artefacts, after which the images are enhanced to highlight the ROI in terms of its border and/or the texture inside its borders. The research presented in this thesis also considers the issue of noise removal for more accurate segmentation of the ROI, which will influence the extraction of effective features and, eventually, more accurate diagnosis results.

Various traditional methods, such as thresholding, region growing, etc., are available for segmenting ROI. As mentioned in Section 1.1, it has well been recognised in the literature that automatic segmentation of ultrasound images is a complex, challenging and domain-specific task, and still an open problem for research. This is precisely the focal point of this research; we attempt to demonstrate that automatic segmentation, when using more sophisticated methods, is feasible and will hence expand the understanding of the topic beyond the scope of the current state-of-the-art.

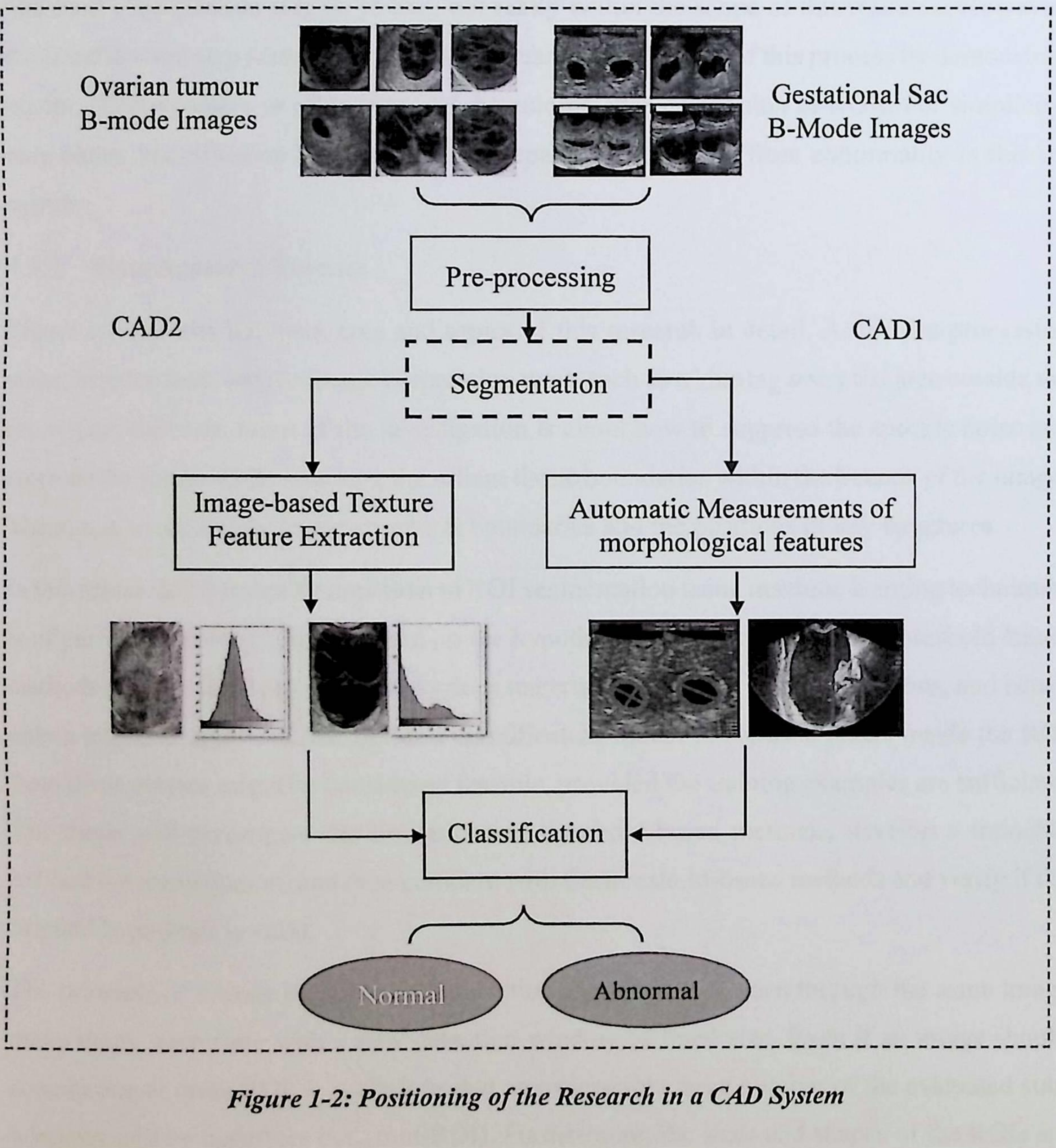



*Figure 1-2: Positioning of the Research in a CAD System*

Both automatic diagnosis, using image-based texture features, and manual diagnosis, using morphological features, rely on the accurate segmentation of the ROI. The more accurate the segmented ROI is, the more precise the features so extracted. Unfortunately, there is no solid ground truth to be used as a reference against which to test the precision of the extracted features, particularly the morphological features of the segmented ROI. The manual measurement results cannot be used as the ground truth because of intra- and inter-observer variations. Although the automatic determination of the segmented ROI must be correlated with an observer's manual measurements, the best way of testing the effectiveness of such measurements is to refer to the accuracy rates in classification.

The final classification step is, in fact, not really within the scope of this research. However, the classification step serves as an excellent means of evaluation of this process by demonstrating the effectiveness – or otherwise – of the automated segmentation of ROIs. For simplicity, only binary classification is considered for separating normality from abnormality in this research.

### 1.3.2 Main Research Focuses

Figure 1-3 outlines the main area and topics of this research in detail. At the pre-processing stage, besides some simple image preparation work such as trimming away the area outside the fan region, the main focus of the investigation is about how to suppress the speckle noise and enhance the image while retaining the salient tissue boundaries within the bounds of the image. The aim is to accurately locate structural boundaries and the positions of any structures.

In this research, the trainable approach to ROI segmentation using machine learning techniques is of particular interest. This is based on the hypothesis that the conventional threshold-based methods may be unable to deal with various uncertainties in ROI boundary regions, and hence only a trainable approach that builds a classification model to separate pixels inside the ROI from those outside might be considered feasible, provided the training examples are sufficient. The thesis will investigate the limitations of threshold-based methods, develop a trainable method for segmentation, and then compare with the threshold-based methods and verify if the original hypothesis is valid.

The principle of a basic trainable ROI detection algorithm is to scan through the same image many times, each time with a new detection window of fixed size. Even if an image should contain one or more ROI, it is obvious that an excessively large number of the evaluated sub-windows will be negatives (i.e., non-ROI). Furthermore, the sizes and shapes of the ROIs are different from one image to another, making it difficult to capture the genuine ROI though the use of a fixed-size window. This research will also investigate trainable object detection techniques to further enhance the performance of the basic trainable segmentation methods.

It must be acknowledged that ROI segmentation can be a complex process. One round of segmentation based on any method may not deliver the correct ROI due to similarities between the actual ROI and certain regions in the background area of the ultrasound image, creating false positive ROIs. This may well be a characteristic that is specific to ultrasound images, and that not observed for other types of medical imaging. Therefore, some forms of post-processing after an initial ROI segmentation may be unavoidable. In this post-processing step, novel solutions need to be developed to reduce the number of false positives found by using additional texture information of both the ROI and the background, by estimating ROI borders more precisely and using geometric properties of ROIs and non-ROIs. We acknowledge the utility in this regard of prior knowledge about the application domains in developing solutions to solve specific problems with the applications. At the same time, we also attempt to limit the use of such knowledge in order to maintain the generality and applicability of the proposed solutions across different application domains.

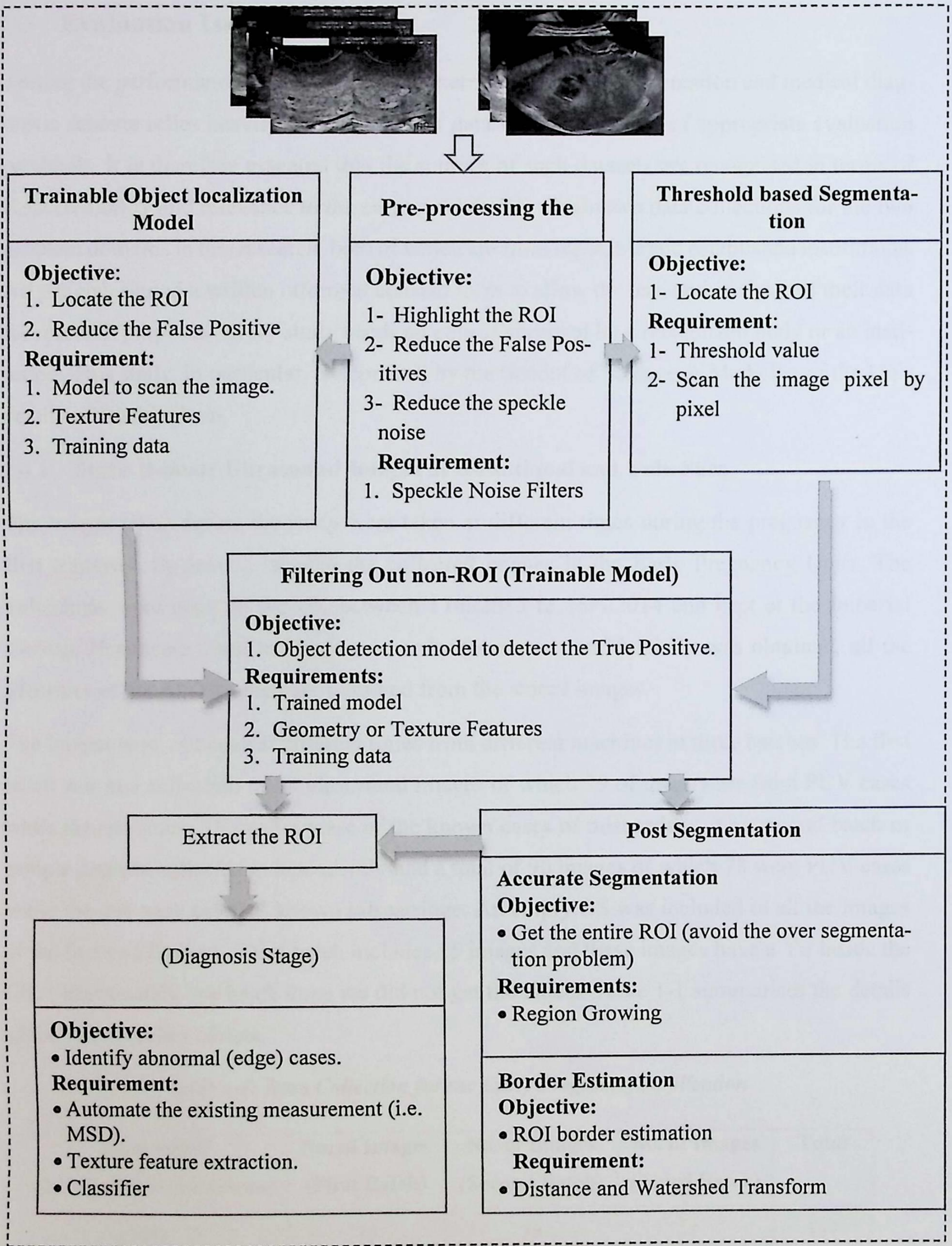



*Figure 1-3: Overview of Main Research Stages*

## 1.4 Evaluation Issues

Testing the performance of any automatic pattern recognition, segmentation and medical diagnostic scheme relies heavily on the choice of dataset(s) and the use of appropriate evaluation protocols. It is therefore essential that the sources of such datasets are recognised in terms of their credibility and relevance to the current work. We obtain two data collections for the two problem domains in this research, both of which are from reputable and established institutions. All patients signed a written informed consent form to allow the use, and analysis of their data for research purposes. Every study needs an ethical approval by a recognized body or an institution. This study, in particular, is approved by the School of Science & Medicine of the University of Buckingham.

### 1.4.1 Static B-mode Ultrasound Images of Gestational and Yolk Sacs

The images taken during the study were taken at different times during the pregnancy in the first trimester. Dr Jessica labelled the collected images in the Early Pregnancy Units. The collections were made in the UK between 17/6/2013 to 16/6/2014 and kept at the Imperial College Healthcare Trust in London. In order to ensure confidentially was obtained, all the identities of the participants were removed from the stored images.

The images were obtained at different times from different machines in three batches. The first batch was in a collection of 94 ultrasound images of which 79 of them were from PUV cases while the remaining 15 images were of the known cases of miscarriage. The second batch of sample datasets collected independently had a total of 90 images of which 78 were PUV cases while the rest were cases of known miscarriage. An empty GS was included in all the images of the first two batches. Third batch includes 15 images and these images have a YS inside the GS. Unfortunately, for batch three we did not get the labels. Table 1-1 summarises the details of the three batches of data.

***Table 1-1: Data Collection for the Early Pregnancy Application***

| Class name (based on Presented Scan) | No. of Images (First Batch) | No. of Images (Second Batch) | No. of Images (Third Batch) | Total |
|---|---|---|---|---|
| PUV | 79 | 78 | - | 157 |
| MC (Miscarriage) | 15 | 12 | - | 27 |
| Total | 94 | 90 | 15 | 199 |

For the images in Batches 1 and 2, each image consists of two views of a GS from the sagittal and transverse planes. All the images in the two batches are provided with GS diameter measurements (d1, d2, d3) and the MSD values in the same Excel file.

Pregnancy was confirmed via a two-stage procedure. Once a pregnancy test produced a positive result, a scan was taken, especially if the pregnant woman reported pain in the abdomen or any bleeding, as well as to check for any irregularities. The results of the ultrasound scans were used to inform the preliminary diagnosis. In the PUV and miscarriage cases sampled in this study, pregnancy was determined based primarily (present diagnosis) on the MSD of the GS. In the case of miscarriage, no additional scan was carried out, but if a miscarriage was not confirmed, further testing was undertaken after a further fourteen days (ultimate diagnosis), including a blood analysis. In our approach, we will use the two steps as ground truth to investigate the effect of the texture features in both cases. Table 1-2 lists the two kinds of labels for the images. It also shows that numerous cases initially classified as PUV were subsequently reclassified as miscarriage cases, but some images do not have the ultimate diagnosis label.

***Table 1-2: Present Scan Diagnosis vs Ultimate Diagnosis***

| Class name | Diagnosis on presented scan | Ultimate diagnosis |
|---|---|---|
| **Not Miscarriage** | 157 | 30 |
| **Miscarriage** | 27 | 101 |
| **Total** | 184 | 131 |

### 1.4.2 Static B-mode Ultrasound Images of Ovarian Tumours

The images used in the study were results of the ultrasound scans conducted on women who participated in the International Ovarian Tumor Analysis (IOTA) studies (Dirk Timmerman 1999). These images were mainly ovarian tumours. The participants of this study were women who had undergone surgical removal of the tumours between November 2005 to November 2013 and had known the histological diagnosis. This was the inclusion criterion used for the study. Ethical considerations were taken into account in seeking permission and this included a written in-formed consent form that allows the collected information to be used for research purposes and data analysis. Permission has also sought the University of Buckingham's School of Science & Medicine. Which granted the researchers and ethical approval to conduct the study.

From the IOTA database, a total of 187 US images were retrieved from the results of 177 anonymous patients (Astraia software gmbh, Germany). This database was accessed at the Gynaecological Ultrasonography Department of Campus Gasthuisberg, KU Leuven, Belgium. In this case, for 10 patients, there were an additional 10 images with another representative ROI. This reflected the final histopathology. The image mainly majored on the surgically removed tumours and presented them in form of a 2D B-mode scan by ultrasound. Among the 187 ultrasound images, 112 of the images were from the benign tumours while the rest were malignant tumours. Each of the used images was selected by Dr Jeoren Kaijser of the KU Leuven Hospital and gave a clear representation of the final histopathology.

## 1.5 Thesis Contributions

The following is a summary of the main contributions achieved in this research to overcome the limitations of existing ultrasound segmentation techniques (in miscarriage and ovarian cases).

1. Developing a new approach to automatically segment the GS from a static B-mode image by exploiting its geometric features for early identification of miscarriage cases. To accurately locate the GS, the proposed solution uses wavelet transform to suppress the speckle noise by eliminating the high-frequency sub-bands and thus prepare enhanced images. This was followed by a trainable segmentation step to locate the GS through the following stages. First, an initial thresholding is used to binarize the image, followed by filtering unwanted objects based on their circularity, size and greyscale mean. A Region Growing technique was then applied in post-processing to finally identify the GS.
2. Surveying and evaluating different speckle noise filtering techniques reported in the literature to enhance ultrasound images such as the Wavelet transform filter, and the Median filter. We illustrate and analyse the effects of the filters on the segmentation and the feature extraction stages.
3. Proposing a novel multi-level trainable segmentation method to achieve three objectives: 1) segmenting and measuring the GS; 2) automatically identifying the stage of pregnancy; and 3) segmenting and measuring the YS. The first-level segmentation employs a trainable segmentation technique based on the histogram of oriented gradients to segment the GS and estimate its size. This is followed by the automatic identification of the stage of pregnancy based on histogram analysis of the content of the segmented GS. The second-level

segmentation is used thereafter to detect the YS and extract its relevant size measurements. A trained neural network classifier is employed to perform the segmentation for both levels.
4. Proposing a new solution to segmenting ovarian masses automatically from ultrasound images. Initially, the method uses a trainable segmentation procedure and a trained neural network classifier to accurately identify the positions of any masses and cysts. The borders of the masses can then be appraised using a watershed transform.
5. Proposing a new Trainable Region Growing model to enhance the traditional RG using texture features and ANN.
6. Developing a hybrid trainable model to segment the ovarian tumour mass by combining the Cascade model with the trainable segmentation method in order to address the over-segmentation problem.
7. Using machine learning as a new abnormality signature to complement the MSD-based decision to identify the miscarriage cases from the PUV. The new signature was based on training commonly used texture features within the segmented ROI to be used as a model for diagnostics.
8. Proposing a new adaptive model to enhance the segmented ovarian ultrasound image and reduce the overlap between the texture feature for benign and malignant tissue. This work has been joint done with Mr.Dhurgham Al-Karawi.

## 1.6 Thesis Layout

The rest of the thesis is organised as follows.

**Chapter 2:** This chapter presents background information about medical imaging, ultrasound scan images, and detailed information about miscarriage and ovarian tumours. In addition, it gives a technical background of different image processing techniques related to this research.

**Chapter 3:** This chapter reviews the existing work in the literature on extracting regions of interest in medical ultrasound images of ovaries in relation to pregnancy and ovarian tumours. Additionally, an evaluation of methods used to evaluate the proposed methods will be given.

**Chapter 4:** In this chapter, we first evaluate existing speckle noise reduction techniques and identify the best filter for both segmentation and identification; traditional segmentation methods will then be evaluated. Based on this evaluation, we propose a new segmentation method that includes a number of stages to extract the ROI. This algorithm has been applied to extract the GS from the ultrasound image.

**Chapter 5:** Effective multi-level segmentation of the gestational and YS using texture-based trainable models to tackle limitations that faced previous algorithms are proposed in chapter 3. These limitations are presented at the beginning of this chapter. Furthermore, we have used the trainable segmentation and watershed transform to identify unilocular and multilocular cysts from ultrasound images of ovarian tumours.

**Chapter 6:** This chapter proposed solution in terms of enhancing the region growing, as based on the neural network classifier and texture features, and justifies why this enhancement is necessary. Then, a method is proposed to improve the trainable segmentation and make it suitable for the ovarian tumour cases that show similarities between the texture of the ROI and the background.

**Chapter 7:** Evaluate the effectiveness of texture-based feature vectors from ultrasound images in identifying miscarriage cases in early pregnancies, as well as propose an adaptive speckle noise model to reduce the speckle noise present in ovarian ultrasound images to enhance the texture feature and reduce the overlap between malignant and benign tissues.

**Chapter 8:** Concludes the thesis with a summary of the major findings of this research and highlights potential future research directions.

# Chapter 2. **Medical and Computational Background**

This chapter is devoted first to describing the basic background of medical imaging and problem domains. It will then briefly introduce various types of ultrasound imaging systems and describe their importance in detecting and diagnosing gynaecological abnormalities and diseases. The chapter will then present an overview of technical knowledge in image processing and machine learning, and provide a general and broad review of the current technological development for the problem domains that are relevant to this research. Overall, this chapter serves as a primer for the understanding of the thesis work. Although, references to relevant literature are provided throughout this chapter to support the specific information provided, more detailed reviews of recent start-of-the-art developments for our specific research objectives will be given in the next chapter and whenever needed in each later key chapter as introductions to our own proposed solutions to each specific problem. Readers who are familiar with the problem domains, medical image processing and machine learning can skip this chapter.

## 2.1 Medical Imaging and Medical Image Analysis: An Overview

Medical imaging refers to technologies that are used to view the human body in order to monitor, diagnose, or treat medical conditions (Bushberg, et al. 2002) (Dhawan 2011). The main goal is to provide an internal picture of bodily structures in a way which is as non-invasive as possible (WiseGeek. 2013). Medical imaging has become one of the most common methods for medical laboratory tests, and has been undergoing a revolution over the past decade that has shown rapid development, greater accuracy, and increasingly less invasive devices (Kasban, El-Bendary and Salama 2015).

The basic concept of a medical imaging system is shown in Figure 2-1. It consists of a source of energy that can penetrate the human body at one end and a sensor that can receive signals at the other. The energy is passed through the body, where it is absorbed or attenuated at differing levels according to, for instance, the densities of the tissues and atomic species it encounters, creating signals that are detected by special detectors or sensors compatible with the energy source. The detected signals are then mathematically manipulated to create an image in two or three dimensions (Kasban, El-Bendary and Salama 2015).

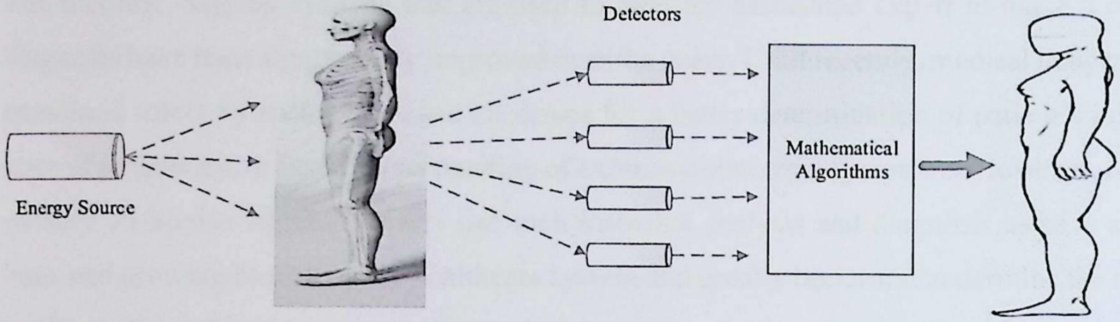



*Figure 2-1: The basic concept of a medical imaging system (Kasban, El-Bendary and Salama 2015).*

Medical imaging is performed in various modalities that are characterised by the source of energy. Commonly used modalities include MRI, CT, and ultrasound, (Sharma and Aggarwal 2010). While each type of imaging modality has its different uses, it is not always the case that one kind of imaging is better than another. Which type of medical image modality should be used depends on the medical condition and diagnosis besides the nature of the examined body organ/tissue. Figure 2-2 illustrates some of the various types of available and widely used medical imaging devices and the images they produce. While some devices such as US and MR capture images of soft tissues and fluids, others such as X-ray depict hard tissue and bone structures. So, all image modalities have their own strengths and limitations. No single medical image modality can deliver all the details required for diagnosis and treatment of clinical conditions (Bushberg, et al. 2002) (Dhawan 2011). Most, if not all, of the image modalities mentioned above are already in practical use in almost all medical centres and hospitals. The growth of such technology has been extensive over the last few decades and has furthered our comprehension of disease aetiology, progression and effective treatment.

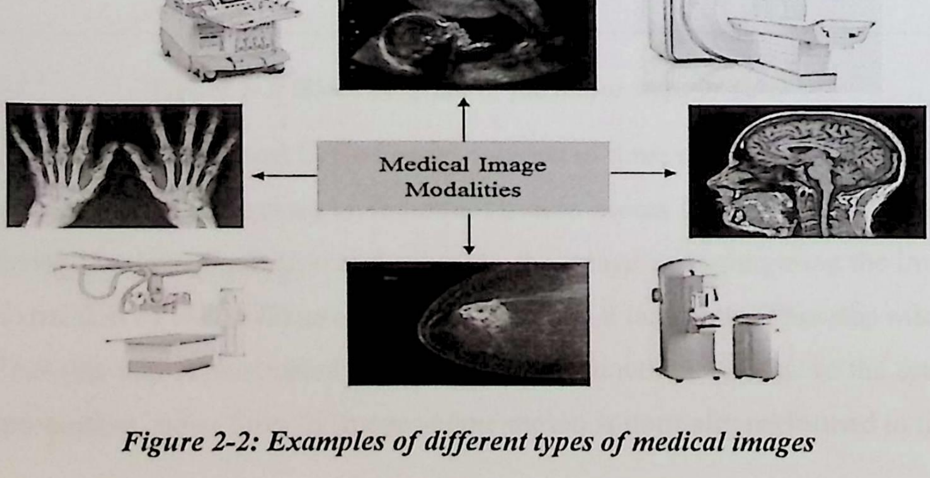



*Figure 2-2: Examples of different types of medical images*

The medical imaging systems that are used to help the associated expert to make a correct diagnosis have been significantly improved over the years. Until recently, medical images were examined solely by radiologists and clinicians for a better determination of patient's injuries, state of pregnancy, or better understanding of exhibited abnormal symptoms. However, relying entirely on human experts to carry out such intensive analysis and diagnosis tasks is adding huge and growing burden on the healthcare system and greatly limits and undermine the ability to efficiently make appropriate, accurate decisions. First, access to well-trained human medical experts with appropriate skills and experience are extremely limited in both developed and developing countries. Appropriate training is extremely costly and takes long times. Second, subjective decision-making may often result in inter– and intra–observer differences in the decision outcomes. Inter–observer differences represent the different outcomes achieved by different observers when scrutinizing the same subject matter, whereas intra–observer differences refer to the different outcomes the same observer makes when scrutinizing the same subject matter repeatedly. Diagnostic errors do not only result in adverse impact on patients health/life but are becoming the subject of growing litigations that result in depleting healthcare systems finances. Both types of inconsistent decision making can be greatly reduced by the adoption of automatic or semi-automatic machine-based systems known as computer-aided diagnosis (CAD) systems. This approach not only help reduce the pressure on healthcare services but can greatly benefit from using advanced computer vision and machine learning technologies (Otoum 2013). Figure 2-3 outlines the major steps of a typical image processing CAD.

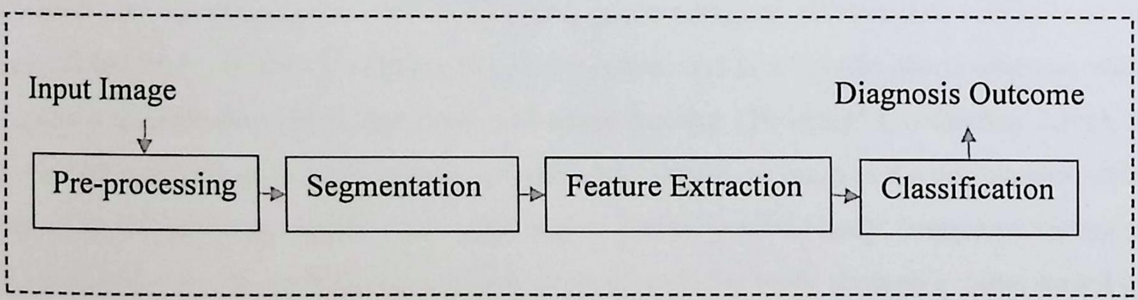



***Figure 2-3: Block diagram of the major steps of CAD***

As shown in Figure 2-3, 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).

## 2.2 Ultrasound Imaging for Gynaecological abnormality detection

Among medical imaging modalities, the ultrasound image is currently one of the most important, widely used, and multipurpose imaging modalities in medicine. The most common use of ultrasound imaging system is in detecting gynaecological abnormalities (Geirsson and Busby-Earle 1991) (Chan and Perlas 2011). An ultrasound is a "cyclic sound pressure with a frequency greater than the upper limit of human hearing (20 kHz)" (Ambedkar 2012). An ultrasound scan is used to create images of soft tissue structures, such as the gallbladder, ovary, liver, kidneys, pancreas, bladder, and other organs and parts of the body. Images on ultrasound are produced using a small probe that is in contact with the body through a water-based gel. This hand-held probe transmits sound waves into the body, and then collects data based on the intensity of the reflected echoes. This data is then interpreted by the machine to display a grayscale image.

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

### 2.2.1 Early Pregnancy and Miscarriage

A regular pregnancy takes 40 (± 2) weeks. The first indicator of pregnancy is the absence of the menstrual period. Pregnancy tests, which are conventionally conducted using a urine sample, capitalise on the presence or absence of human chorionic gonadotropin (HCG) (Doubilet, et al. 2013). In situations where pregnancy tests of this kind yield positive results, an initial scan may be conducted to identify the GS's age and, based on this, to calculate the birth date. As can be seen in Figure 2-4, the structural anatomical elements of early pregnancy are as follows: (i) the GS is the first visible structure in the very early stages of pregnancy (Figure 2-4-B); and (ii) the yolk sac is a ring-shaped structure, which is the first structure to be identified within the GS (see Figure 2-4-C) (Nebraska 2013). An embryo will be seen along with the yolk sac, which will develop foetal heart activity. Failure to identify foetal heartbeat in an embryo that measures more than 7 mm in length is diagnostic of miscarriage (Nebraska 2013). Figure 2-4-D shows the embryo attached to the YS inside the GS. The initial structural element that is observable inside the GS is the YS. At the point where the MSD of the GS is 5-6 mm, the YS can often be observed by employing trans-vaginal ultrasound (UIDELINE 2005).

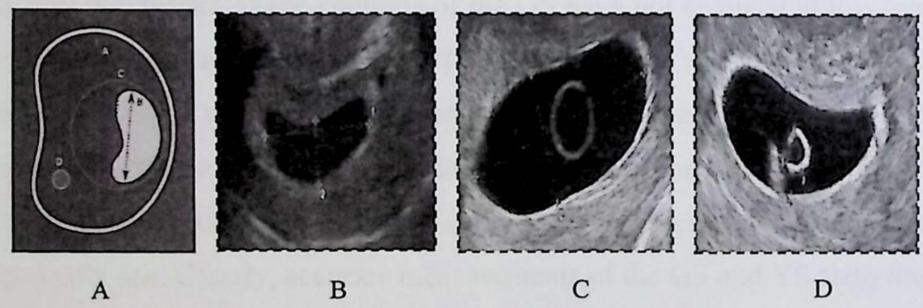



*Figure 2-4: Examples of ultrasound images taken at the very beginning of a pregnancy until developing the embryo (A) The anatomical structures of the early pregnancy: A) Gestational sac, (GS), B) Crown-rump length (CRL) of the embryo, C) Amniotic sac and D) Yolk sac*

Statistical evidence reveals that a range of complications can occur during pregnancy, the most prevalent of which is a miscarriage. In the UK, miscarriages figure reach nearly a quarter of a million each year (Khazendar, Al-Assam, et al. 2014) (Khazendar, Farren, et al. 2015). As reported in (Giakoumelou, et al. 2015), around 20% of all pregnancies are miscarried prior to 24 weeks, and the majority of these occur in the first trimester (namely, within 12 weeks of conception).

Gestational assessments conducted over the course of the first-trimester focus on the confirmation of fundamental details regarding the pregnancy, including its location, viability, and whether it is a twin (or higher order) pregnancy. An assessment of the dimensions of the GS and the YS provide information about the probable gestational age of early pregnancy, where such assessments are also valuable in diagnosing miscarriage (Preisler, et al. 2015). The mean sac diameter (MSD) is generated from three diameters measured in the sagittal and transverse planes (Ectopic pregnancy and miscarriage: diagnosis and initial management 2012). As the pregnancy develops, the YS and foetus form, and their size and growth, along with the growth of the GS, become indicators of wellbeing.

Different studies have considered that miscarriage should be declared based on different cut-off values for MSD within the range of 13-25 mm (Bourne 2016) (Levi, Lyons and Lindsay 1990). As displayed below, the most up-to-date limits for miscarriage diagnosis can be defined as (miscarriage identification cut-offs according to the NICE guidelines):

- Mean gestational sac diameter (MSD) of ≥ 25 mm with no obvious foetal pole.
- Mean sac diameter of (GS) of ≥ 25 and no embryo defined in YS.

In cases where an MSD of an empty GS greater than 25 mm is observed, a diagnosis of miscarriage can be safely made (Preisler, et al. 2015). If the MSD is less than 25 mm, a scan must be repeated after two weeks. If the contents of the GS have not changed at this repeat scan, a diagnosis of miscarriage can also be made. If a GS contains a YS and/or measurable foetus, but foetal heart activity is not present, a repeat scan is recommended after one to two weeks before a diagnosis of miscarriage can be made. Indicators (but not diagnostic criteria) of miscarriage include a smaller GS (or foetus) than anticipated, as based on the last menstrual period or an enlarged yolk sac. Clearly, accurate measurements of the GS and YS sizes are vital.

### 2.2.2 Ovarian Cancer

Cancer starts when the cells in a part of the human body start to multiply in an out of control manner. There are more than 100 types of cancer, including ovarian cancer (Cancer stats key facts 2011). Ovarian cancer specifically refers to cancer cells starting from the ovary (cancer cells from other parts of the body can also invade the ovary but are not considered to be ovarian in nature). Symptoms of ovarian carcinoma do not appear until an advanced stage, and therefore ovarian cancer is known as the "silent killer". Over 7,400 women in the UK are diagnosed with ovarian cancer each year (Cancer Research UK 2015). Ovarian cancer in the UK is the sixth most common carcinoma in females after bowel, breast, lung and womb cancer (Cancer Research UK 2015) (American cancer society 2014).

Ovarian cancer is one of the more common forms of cancer, particularly in the post-menopause population of women, and has the highest mortality rate of all gynaecologic cancers (Jeong, Outwater and Kang 2000) (Fishman, et al. 2005). To date, the lack of specific associated symptoms has been the leading cause of death (Berek and Bast Jr 2003). The most important factor in treating this 'silent killer' is to accurately characterise and determine the state, the stage and even the type of ovarian tumour present. Correct classification is necessary to prevent unnecessary procedures, such as surgery, for those patients with benign masses, and to follow an optimal treatment route for those with malignant masses.

Ultrasound imaging has become the most widely used method to identify and distinguish tumour state (i.e., malignant from benign) and distinguish tumour types. A B-mode ultrasound imaging can depict morphological features including unilocular or multilocular cysts, presence of fluid, solid tissues, internal wall structures, papillary projections and acoustic shadows. A superimposed Doppler image on top of a B-mode image reveals blood flow information within tumour areas, offering additional assistance in tumour diagnosis. The results of combining these different types of information enable clinicians to determine the seriousness of the imaged mass (Sayasneh, et al. 2015) (Sayasneh, et al. 2016). To gain a more accurate classification of tumour types, researchers from the International Ovarian Tumour Analysis (IOTA) group have created a series of model systems utilising sonography and other patient information including Risk of Malignancy Index (RMI), Simple Rules, Logistic Regression models (LR1 and LR2), and the most recent ADNEX risk model (Van Calster, et al. 2014). However, all these models have their own, different levels of accuracy, and which are still considered to be inferior to the expert opinion and decisions of the appropriate domain expert (Kaijser 2015).,

The Assessment of Different NEoplasias in the adneXa (ADNEX) model has recently been developed by the IOTA group scientists. This model is able to differentiate between early and late stage (II-IV) primary cancers, secondary metastatic cancers, borderline tumours and benign tumours. The ADNEX model is based on six ultrasound parameters and three clinical parameters that offer risk calculation with and without CA125. The six ultrasound parameters used in the ADNEX model are lesion diameter in mm, solid tissue proportion, number of papillary projections, number of cysts (10 or more, yes or no), the presence or absence of acoustic shadows and the presence or absence of ascites (Van Calster, et al. 2014). Validation of the model was achieved using parameters collected by experienced ultrasound clinicians with a special interest in gynaecological ultrasonography, and who were equivalent in experience and knowledge to UK radiology consultants (Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB)). The benefit of this model is that it allows for effective patient triage due to its ability to identify tumour type. As such, patients can be rapidly assessed and assigned to the most suitable management pathways for their tumour type, whether that is surgery, conservative follow-up or treatment at a specialist cancer unit. However, it is also beneficial to be able to identify the tumour malignancy sub-type as less aggressive treatment options are available for early-stage ovarian cancers and borderline ovarian tumours, the significance being that less aggressive treatment often allows younger women to retain their fertility. Identification of metastatic ovarian cancers should, nevertheless, be treated in the same manner as primary cancer (Van Calster, et al. 2014) (Sayasneh, et al. 2016).

To date, the image parameters that are vital to the diagnosis models for both miscarriage and ovarian cancer diagnoses, working properly are extracted from the ultrasound image manually by the gynaecologists. The inter- and intra-variations of the manual measurements can be a source of misdiagnosis with the obvious dire consequences. Therefore, the objective of this research is to locate the ROI and extract the measurements for those particular parameters automatically.

## 2.3 Computational Background

Medical images can be segmented and annotated either manually or automatically to characterise specific parts of the images. Although different medical image modalities encapsulate different features and have different objectives, they share some common problems in their practical use. In general, medical images may lack visual clarity and contain different types of noise. To extract useful information from such images, adequate image processing techniques are needed. In the following subsections, a systematic review of the main image processing techniques related to this research is given.

### 2.3.1 Pre-processing (De-noising)

There are various types of noise known to degrade images including: Gaussian noise, frequency noise, impulsive noise and multiplicative noise (Mateo and Fernandez-Caballero 2009). In ultrasound images, the noise content is multiplicative and non-Gaussian in nature. A model of multiplicative noise (Wu, Zhu and Xie 2013) (C. P. Loizou 2005) is given by:

$$J(i,j) = I(i,j) * N(i,j) \qquad (2\text{-}1)$$

where the noisy image $J$ is the product of the original image $I$ and the non-Gaussian Noise $N$, where (i , j) represents the location of an image pixel.

Ultrasound imaging suffer from a specific noise type known as speckle noise, which is somewhat different from the above reasonably well understood noise types. It is due to reflection and scattering of sound waves within the tissue (see Figure 2-5). In this figure, we can see clearly the effect of the speckle noise on an object of interest. Speckle noise degrades tissue texture, decreases the visibility of small amounts of contrast and resolution, and makes continuous structures appear discontinuous (C. P. Loizou 2005). Speckle noise is generally more difficult to remove than additive noise because the intensity of the noise varies with pixel intensity in the image. Therefore, it is primary factors such as the above that limit the effective application of image processing and image analysis. Accordingly, speckle de-noising will be investigated in this thesis.

The main purposes of reducing speckle noise in the medical ultrasound image are to improve the contrast between objects within the image with clearer borders such that the human interpretation of the image content becomes easier. Speckle noise reduction can also serve the purpose of pre-processing the image for automatic image processing tasks such as image segmentation for ROIs (Tamilkudimagal and Kalpana 2011).

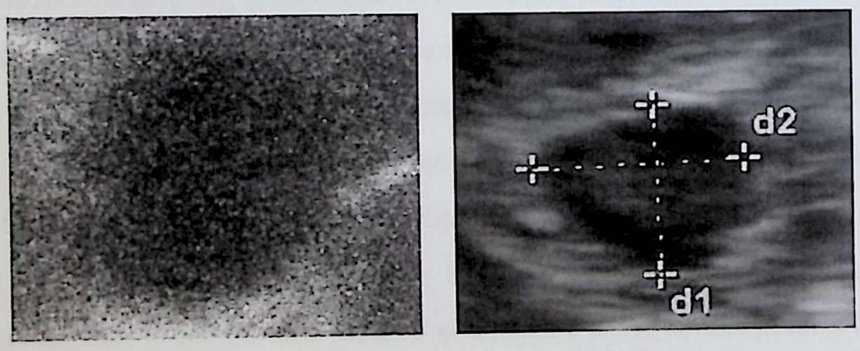



*Figure 2-5: Illustrate the effect of speckle noise*

In general, noise removal is a challenging task that is dealt with using a variety of spatial and frequency domain filtering. In the spatial domain, statistics order filters (e.g. mean, median, maximum, minimum, adaptive median, Harmonic and geometric means) are commonly used to deal with some or all of the above types of noise, but the problem remain a challenge in the case of severe level of noise. In fact, experience reveals that Speckle noise removal is far more challenges for such approaches.

Speckle noise in ultrasound images had been studied since the 1970s when investigators described its statistical and fundamental properties (Abrahim, et al. 2012). In the literature, several methods have been used to reduce the effects of speckle noise on ultrasound images (Sudha, Suresh and Sukanesh 2009) (Karthikeyan and Chandrasekar 2011) (Mateo and Fernandez-Caballero 2009) (Abrahim, et al. 2012) (Rabbani, et al. 2008) (Gai, et al. 2018) (Garg and Khandelwal 2018). All these studies were based on two approaches currently being adopted within research into speckle noise reduction. One speckle filtering method consists of a window (N*N) to cover each pixel with the neighbourhood and applies a mathematical calculation to determine the value of the central pixel; this process moves the image over each pixel in the image, whilst the other uses the entire image, i.e., the Fourier transform filter works through its application to the whole image. Many of the techniques that can be found in the existing literature describing speckle noise reduction involve the use of filters such as the median filter, Wavelet transform filter, Fourier transform, etc. We shall now describe and summarise the use of these filters.

**The Median Filter**

Median filter is a nonlinear filter that is used in the spatial domain of an image. The applicability of medial filter has been recognized due to its significant value in supressing the impulsive noise. The medial filtering is used as an effective pot-filtering technique to reduce speckles. This filtering technique applies a median intensity with suitable sized and shaped region denoted by $W_{ij}$ surrounding the pixel $(i, j)$ of interest, which acts as the output pixel value. This application assists in eliminating the presence of impulsive artefacts in the pixels, which is below half of the region size $||W_{ij}||$ (Chen, Broschat and Flynn 1996). (Loupas, McDicken and Allan 1989) has proposed a new technique that can ease the speckle noise. An adaptive weighted median filter is used in this method and is used based on the pixel's weighting coefficient in the window. The value of weighting coefficient in appearing many times as the weight when estimating the median affects every pixel. Thus, if there are similar weights, this method acts in a similar way like the typical median filter. However, when the weights are not the same, the median decreases from the centre of the window to the outer limits and the details as well as image edges are less altered. Similarly, there will be elimination of less noise when using the method. Therefore, compromising the image preserving and supressing noise reflects weight choice. As a result, the characteristics of an image within the window allow for the algorithm to adapt the weights and this occurs for every step involved in image processing. The local statistical analysis thus determines the image characteristics (Loupas, McDicken and

Allan 1989) (Mateo and Fernandez-Caballero 2009) (Anqing 2010) (Mateo and Fernandez-Caballero 2009).

**The Wavelet Transform**

An example of frequency domain transformations is the Wavelet transform (WT). WT represents an input signal as a family of orthonormal wavelet bases on a single wavelet function, referred to as the mother wavelet, through repetitive translations. WT involves using filters in the transformation. The most notable and widely used cases are the Daubechies (DB) family of filters such as db2 (also known as the Haar filter), db4, db6, and db8 of length 2, 4, 6 and 8, respectively (AL-Hassan 2014). Using wavelet and scaling functions, the WT method can break down an image into four different frequency bands. The first step of the decomposition stage is where the input image is decomposed into four different sub-bands, called the HH (high-high), HL(high-low), LH (low-high) and LL (low-low) sub-bands. Each sub-band shows the information of the image on a different plane. For example, the HH shows diagonal information, HL shows the horizontal information, LH sub-band shows the vertical information, and the LL provides the low-frequency information. The next step of the decomposition process includes further division of the LL sub-band (Kumar, Dutta and Lehana 2013).

WT has been successfully used for image processing and analysis, including reducing the ultrasound speckling effect. In fact, WT is the preferred method for reducing speckle as it can efficiently identify and isolate redundant information. There are two basic approaches currently being adopted in research into the speckle noise reduction using WT:

- Eliminating certain high frequencies. When an input ultrasound image is decomposed by using WT, the LH, HL and HH sub-bands contain the image's high frequencies. Setting the coefficients within the sub-bands to zero when the image is reconstructed by performing the inverse of the wavelet transform can reduce the speckle noise effect. However, reducing the high frequency in such a way does not mean that all speckle noise in the ultrasound image is eliminated, and some of the information of the ultrasound image may also be lost (Mateo and Fernandez-Caballero 2009).
- Eliminating certain high frequencies at a threshold. This type of technique is similar to the first type, but instead of totally removing high frequency elements, a threshold for each frequency sub-band is computed. Only the frequencies above the threshold are removed in the inverse process of reconstruction. This technique has been applied in

various studies such as (C. P. Loizou, et al. 2005) (Karthikeyan and Chandrasekar 2011) and (Rabbani, et al. 2008).

**Fourier Filtering**

It is a type of filtering that is mainly based on the theories of Fourier transform. It is associated with the analyses of a signal into its spectral frequency elements. It converts the signal from time or spatial domain into a frequency domain. During transformation, the sum of sines and cosines of multiple frequencies multiplied by a different coffeicient is expressed as any function that periodically repeats itself. This is termed as the Fourier series (we would now be able to call this aggregate a Fourier series) (Mateo and Fernandez-Caballero 2009). The most important characteristics of these representation is that can allow us to return or recovered the information from Fourier domain to the original domain without losing information by taking the inverse process. The main purpose of the Fourier transform filter is to reduce and minimize any high-frequency elements. We will consider Butterworth and Homomorphic filter as a Fourier filter. In this work, Butterworth filter will be used as the Fourier filter and will be used to smooth this cut off and remove Gibbs effect and the excessive reduction of high frequency image details without losing noise elimination capacity.

The concept of the Homomorphic filter is to change multiplicative noise to additive noise, trailed by a utilization of a low-pass filter, which intends to lessen the additive noise. The homomorphic filter is regularly utilized as it is anything but difficult to execute and generally viable. Homomorphic filtering is accomplished by figuring the FFT of the logarithmically compacted image, utilizing a homomorphic filter to de-spot the clamour and lastly, applying the reverse FFT to make the de-speckled image (C. P. Loizou, et al. 2005) (Mateo and Fernandez-Caballero 2009).

(Mateo and Fernandez-Caballero 2009) has compared the speckle noise algorithms that we have explained above. US image used in the comparison is one for a kidney. From the US image, they generated the noisy image by applying the speckle noise filter on the noisy image. They then compared the results obtained with quality ones produced from the Fourier filter. With these results, they argued that the best quality images are only obtained through Fourier filters. Although the other filters have been found to be able to improve in quality, they have been found not to produce to the quality of the Fourier filter. The use of Fourier filter is also simple compared to other filters since it only takes in one parameter. Wavelet filtering has also not been recommended by many studies to be used in US images. This is because wavelet filtering does not eliminate a lot of noise. Wavelet also eliminates bands that make the white dots appear hence distorting the quality of the image. This shows that homomorphic filters can only be used with the Fourier filter but not useful when used with other filters.

**Wiener filter**

This is a linear filter that is mainly applied adaptively to an image. It is mainly tailored to the local image variance. The filter does a little smoothing when the variance is large and performs more smoothing where the variance is small (Kaur and Singh 2010) (Patidar, et al. 2010). This approach has been found to produce more improved results than the linear filters (Kaur and Singh 2010). The use of this filter preserves the edges and other parts of images with high frequency since it is more selective when compared to the linear filters. The only drawback of the Wiener filter when compared to the linear filter is that it requires more computational power.

Most of the previous work has been aimed at reducing the amount of speckle noise in the US image. Reducing the speckle noise may be very important but it is not guaranteed for the production of accurate image segments. This is because speckle noise is not the only factor that affects the accuracy of image segmentation. One of the objectives of this work is, therefore, to evaluate different filters and recommend one that can reduce speckle noise and at the same time highlight the ROI for the image segmentation task.

### 2.3.2 Image Segmentation

Image quality mainly affects the US image segmentation. A low-quality image will not be able to show all the properties of the image that are necessary for a successful segmentation. The components that make the image produce low-quality include; shadows, speckles, attenuation and drop out of signals of the image. This may lead to loss of image boundaries due to missing boundaries due to the orientation dependence of acquisition.

More problems come in due to contrast. The major areas of interest in the image tend to have low contrast. The main methods of image segmentation that have been applied in the contemporary medical practices include; region-based, watershed and thresholding methods and have proven to be very effective (Stolojescu-crisan and Holban 2013). All these methods exclusively depend on the intensity of the pixels that make an image in the segmentation process. Another major method that has also been proven to be very effective in image segmentation is the use of a Hough transform. This method mainly uses a fixed shape matching feature that enables it to extract and exploit the required segment of the image object from a particular shape. More recently, machine learning based methods offer an alternative for segmentation by classifying pixels as either inside or outside the ROI. These types of segmentation methods are reviewed in the following section.

**Thresholding**

Thresholding is an image segmentation method that is used to segment scalar images. This method creates a binary partition based on the intensities of the image. First, the value of a particular intensity is determined and is called a threshold. Secondly, the pixels with intensities that are greater than the threshold are grouped in one class and those with lower intensity than the threshold are grouped in one class. Figure 2-6 shows two potential thresholds at the valleys of the image histogram. The process of determining more than one threshold is called multi-thresholding.

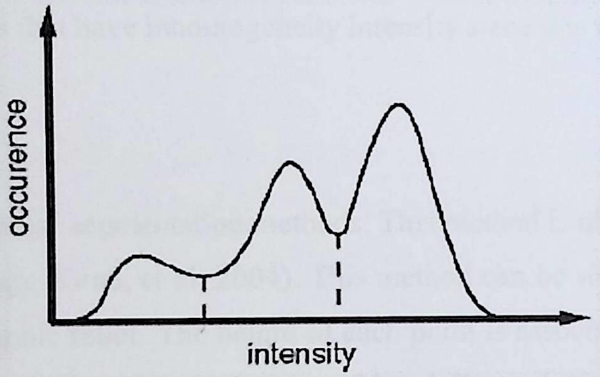



*Figure 2-6: Histogram showing three apparent classes*

This process is mainly applicable in images whose pixel intensities vary from one structure to the other. Although automated methods exit, the partitions can be usually generated interactively. The very well-known methods of thresholding include the Otsu method (N. Otsu 1975) and the use of maximum entropy ( Kapur, Sahoo and Wong 1985).

The method that uses threshold mainly suits images whose structures have different pixel intensities or those whose contrast of the structures vary greatly with the background of those structures. The process of selecting a threshold to use is usually difficult in cases where the noise is too much or when the contrast of structures and the background is too low. Generally, the threshold-based segmentation method was employed as an initial segmentation stage, since it does not give the desired results in the segmentation process. (Fasihi and Mikhael 2016).

**Region Growing**

This is a method used to extract a particular region of the image and is mainly done based on a clearly redefined stated criterion. The criteria that are mostly used include; the intensity of the information and or the edges of the image. In general, this method requires the seed point to be manually selected by an operator. The next step involves extracting the pixels that are connected to the used criterion in some way (Rangayyan 2005). One common criterion may be to grow a given region to the point where the edge of that image is encountered.

There are a number of disadvantages associated with the region growing method of image segmentation. First, using automatic methods of selecting the point to act as the seed and the number of seeds provided by the user can sometimes not be enough to assign each pixel to a region. Secondly, providing the seeds incorrectly such that two seeds that should be on the same region are incorrectly provided in the algorithm, then two distinct points will be created instead of one that is required. Lastly, it has been found to be very challenging to carefully select a criterion such as the texture of a feature or intensity in order to avoid under-segmentation. This challenge is more intense with images that have inhomogeneity intensity since it is very hard to select the suitable threshold.

**Watershed Transform**

This is one of the most popular segmentation methods. This method is obtained from the field of mathematical morphology (Grau, et al. 2004). This method can be simply described intuitively in form of a topographic relief. The height of each point is associated with a grey level then rain that is gradually falling on the terrain is considered. Watersheds are taken as the lines that separate the "lakes" or the catchment basins that form at the edges. The gradient of the original image is used to calculate the watershed transform in this case. The water catchment areas are, therefore, located on the high gradient points.

This method has been applied in the medical field as well as other many areas where image processing services are required. The method has been widely applied due to a number of advantages over other methods of image segmentation. The major advantage is that it is a simple intuitive method. It is also fast and can be parallelized to produce complete image segmentations even when the contrast between the structures and the background is very low. This limits the need for contour joining.

Another advantage is that many researchers have proposed the use of this technique with the multistate framework hence provide the advantages of these representations. There are, however, a number of serious disadvantages associated with this method. The major ones include; over-segmentation associated with the method. Also, this method of image segmentation is very sensitive to noise since it works by finding the minimum values of the image which are majorly affected by noise. If noise affects a number of these minimums, it will lead to over-segmentation (Straka, et al. 2003).

**Hough Transform**

Localisation of geometrical forms in an image can be achieved with the common transform called the Hough Transform (HT) (V C 1962) (Nixon and Alberto 2012), which is especially effective in identifying lines, circles and ellipses in binary images. The HT works by plotting the points of an image from geometric space in accumulator space known as Hough space. An explanation of circle identification via HT is provided in the following part, since the present study puts great emphasis on circle identification.

The formula for characterisation of a circle form is:

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \qquad (2\text{-}2)$$

The above formula indicates where $(x, y)$ is situated on a circle of radius $r$ and centre point ($x_0$, $y_0$). There are two options for visualisation of this formula, namely, as a locus of points $(x, y)$ or as a locus of points $(x_0, y_0)$ with the centre on $(x, y)$ with radius $r$, as shown in (Figure 2-7-A) and (Figure 2-7-B), respectively.

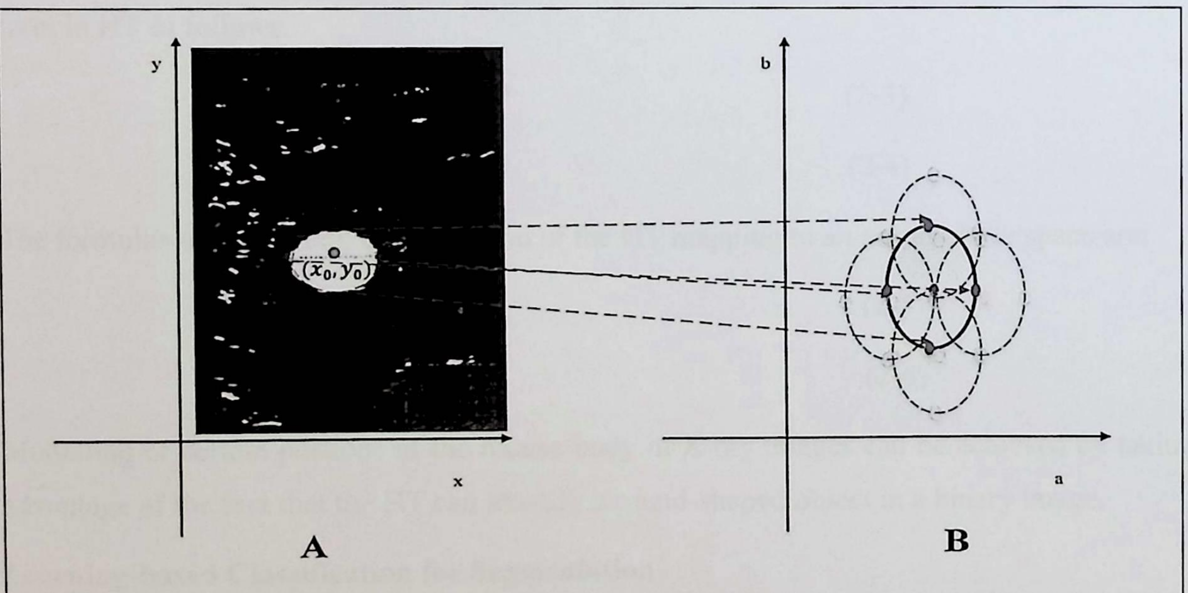



*Figure 2-7: A) Circle located on the parameter plane ($x_0$, $y_0$) and B) transform of the circle to the a and b space*

The latter option implies that every edge point on the circle could serve as the middle of a new circle of radius $r$. A three-dimensional accumulator space is obtained by expanding the characterisation for distinct values of $r$ for every edge point (see Figure 2-8). The circle parameters in the initial image determine the highest number of circle-derived votes produced in an accumulator space.

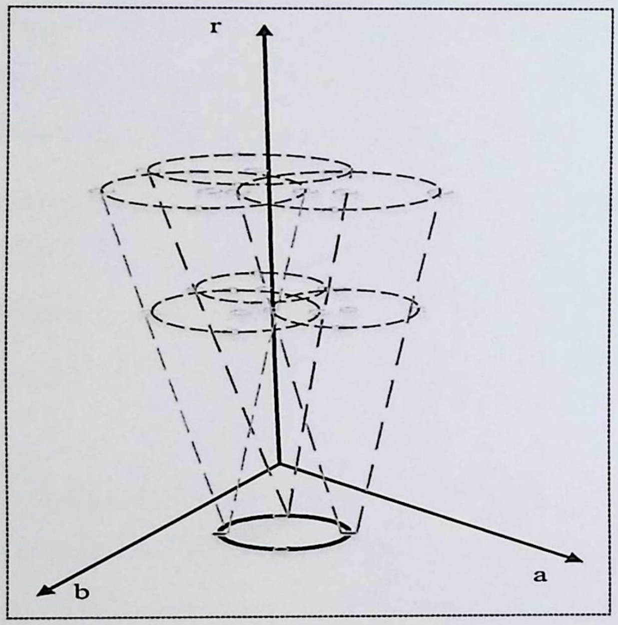



*Figure 2-8: 3D discrete Hough accumulator space (Nixon and Alberto 2012)*

Formulation of a solution for the circle parameters requires the use of the circle parametric form in HT as follows.

$$x = x_0 + r\cos\theta \tag{2-3}$$

$$y = y_0 + r\sin\theta \tag{2-4}$$

The formulae underpinning the definition of the HT mapping to an accumulator space are:

$$x_0 = x - r\cos\theta \tag{2-5}$$

$$y_0 = y - r\sin\theta \tag{2-6}$$

Modelling of certain portions of the mouse body in X-ray images can be achieved by taking advantage of the fact that the HT can identify a round-shaped object in a binary image.

**Learning-based Classification for Segmentation**

This is a pattern that seeks to partition a feature space derived from the image. It mainly recognizes the patterns using data based on known labels (Bezdek, Hall and Clarke 1993). Range space of any function of the image is called the feature space. The feature with the most intensity is called the are the image intensities themselves. The pixels that have their features associated with each other are grouped together until a n-class is obtained (Pham, Xu and Prince 2000).

The training data are manually segmented and are then used as references in automatically segmenting new data. Because of this, the classifiers are sometimes referred to as supervised methods. One simple classifier is a nearest-neighbour classifier where each pixel is classified in the same class as the training data that have the nearest intensity.

Distinct quantifiable features need to be segmented in cases of standard classifiers. Since the training data can be labelled, classifiers can transfer these labels to new data provided that feature space can sufficiently distinguish each label. Unlike threshold methods, no iterative classifiers are efficient in terms of computation and can, therefore, be applied in multichannel images.

The major disadvantage associated with classifiers is that spatial modelling cannot be performed. Recent work on this method has, however, addressed this drawback. In the recent work, the classifier methods have been extended to be able to segment intensity in-homogeneities - corrupted images (Wells, et al. 1996). The other disadvantage arises from the need to use manual interaction methods in obtaining training data. Sets of training data can be obtained from the particular images that require to be segmented although this can be tedious and time consuming. The use of the same training data sets, on the other hand, can lead to biased results especially when a large data set from a number of scans is used. These biases come in since it does not take into account any anatomical and physiological variability between different subjects (Pham, Xu and Prince 2000).

### 2.3.3 Texture Feature Extraction

Method of extracting features helps in simplifying the amount of data necessary to completely describe a huge set of data in a more efficient way. In this case, a few details will only be required to comfortably describe a huge set of data. In order to analyse a set of complex data, the major issue is to determine the number of variables involved in the problem at hand. The problem with a large number of variables is that it will require huge computational power to process and a large amount of memory. This may require a classification algorithm that over fits training sample used and hence can lead to new poorly generalized samples.

Extraction of a feature is a method of getting rid of such problems and still being able to describe data in the most sufficient and accurate way. It mainly uses the visual and texture characteristics of the surface (object under the study). The use of texture in analysis helps in determining the unique characteristics of a texture that can be used to describe that surface in the most simple but unique form. The unique characteristics obtained can then be used to accurately segment and classify objects. The use of texture plays a huge role in image analysis and recognizing patterns but there are still limited architectures that are able to implement the onboard extraction of texture features (Mohanaiah, Sathyanarayana and GuruKumar 2013). Texture features that will be used in this research to address two objectives: 1) segmentation; and 2) identify the abnormality in the image. There are a large number of techniques that have been used for feature extraction based on image texture in a variety of medical image analysis and these include: the Grey-Level Co-Occurrence Matrix (GLCM), Histograms of Oriented Gradients (HOG), Local Binary Pattern (LBP), and Fractal Dimension (FD). We shall now describe each of these types of features.

**Grey-Level Co-Occurrence Matrix (GLCM)**

Originally developed for classification of satellite and aerial images, the Grey-Level Co-Occurrence Matrix (GLCM) technique was among the first to be employed to analyse image texture, and is still the main technique for texture-based feature extraction. A co-occurrence matrix is the joint probability occurrence of grey levels $i$ and $j$ for two pixels with a defined spatial relationship in an image. The spatial correlation between the intensity of two pixels is determined based on detection of a pixel pair $(i, j)$ separated by a distance $(d)$ in an established direction angle $(\theta)$ (Sharma and Singh 2001). Each entry in the GLCM corresponds to a certain grey level arrangement. Provided that every possible pair of pixels has a specific distance in four distinct directions, this technique allows the creation of four distinct matrices. Figure 2-9-A and Figure 2-9-B illustrate the GLCM matrices and their formation (Albregtsen and others 2008) (Mohanaiah, Sathyanarayana and GuruKumar 2013).

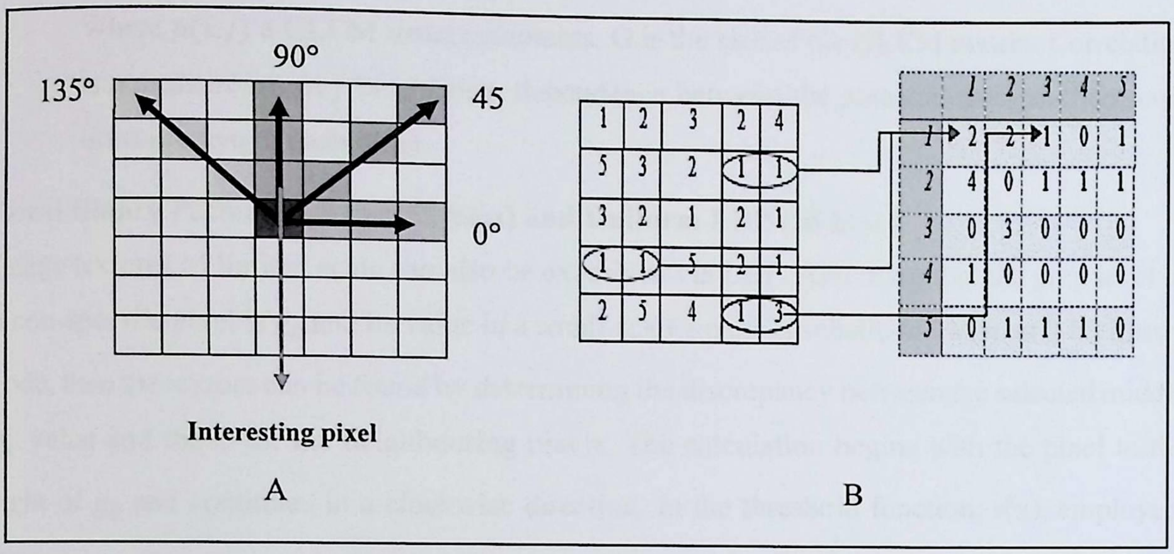



*Figure 2-9: GLCM process: A) Spatial relationships of the image pixels in different angles and B) an example illustrates the formation of a co-occurrence matrix from the grey scale image.*

Several possible texture features may be extracted from a GLCM. The following are summarised in (Albregtsen and others 2008):

- **Homogeneity, Angular Second Moment (ASM):**

$$ASM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\{p(i,j)\}^2 \quad (2\text{-}7)$$

G present the size of the matrix (GLCM matrix). $p(i,j)$ is a GLCM matrix elements. ASM is a feature that present the homogeneity of an image. A homogeneous image will contain only a few grey levels, giving a GLCM with only a few but relatively high values of $p(i,j)$. Thus, the sum of squares will be high.

- **Contrast**

$$CONTRAST = \sum_{n=0}^{G-1} n^2\{\sum_{i=1}^{G}\sum_{j=1}^{G} P(i,j)\},\ |i-j|=n \quad (2\text{-}8)$$

Where G is the size of the GLCM matrix. This measure of contrast or local intensity variation will favour contributions from $p(i,j)$ away from the diagonal, i.e. $i \neq j$.

- **Entropy**

$$ENTROPY = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} p(i,j) \times \log(p(i,j)) \quad (2\text{-}9)$$

Where $p(i,j)$ a GLCM matrix elements. G is the size of the GLCM matrix. Inhomogeneous scenes have low first order entropy, while a homogeneous scene has high entropy.

- **Correlation**

$$CORRELATION = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{\{i,j\}\times p(i,j)-\{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \quad (2\text{-}10)$$

Where $p(i,j)$ a GLCM matrix elements. G is the size of the GLCM matrix. Correlation is a measure of grey level linear dependence between the pixels at the specified positions relative to each other.

**Local Binary Pattern (LBP) (255 bins) and Uniform LBP (58 bins)**

Image textures of limited scale can also be examined via LBP. Assuming that the grey level of a non-specific pixel is $g_c$ and its value in a small-scale image is substituted with an 8-bit binary code, then the texture can be found by determining the discrepancy between the selected middle $g_c$ value and those for the neighbouring pixels. The calculation begins with the pixel to the right of $g_c$ and continues in a clockwise direction. In the threshold function, s(x), employed, difference values that are lower or higher than zero are given values of zero or one, respectively. The subsequent step involves the creation of the LBP image through the conversion of the binary codes of zeroes and ones into decimal values (Ojala, Pietikäinen and Harwood, A comparative study of texture measures with classification based on featured distributions 1996) (Ojalaa, Pietikainen and Maenpaa 2002) (Pietikainen, et al. 2011). The LBP procedure is presented in Figure 2-10. The mathematical expression of this feature takes the following form:

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \qquad S(x) = \begin{cases} 1 & si\ x \geq 0 \\ 0 & otherwise \end{cases} \tag{2-11}$$

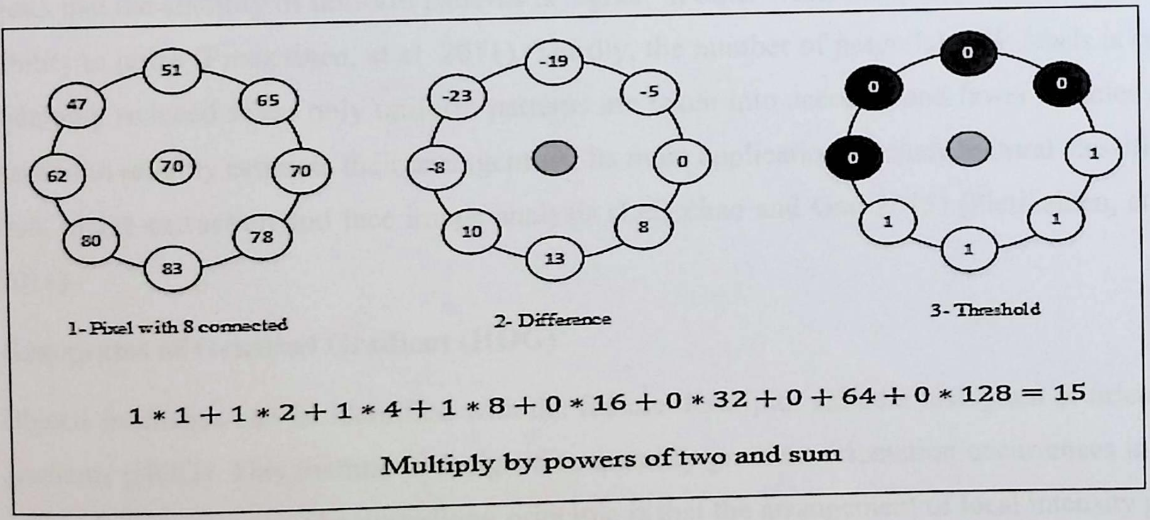



*Figure 2-10: LBP Example (Pietikainen, et al. 2011)*

The local spatial structure of the image is defined by LBP, which presents the qualities of grey invariance and rotation invariance.

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 2-11), 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).

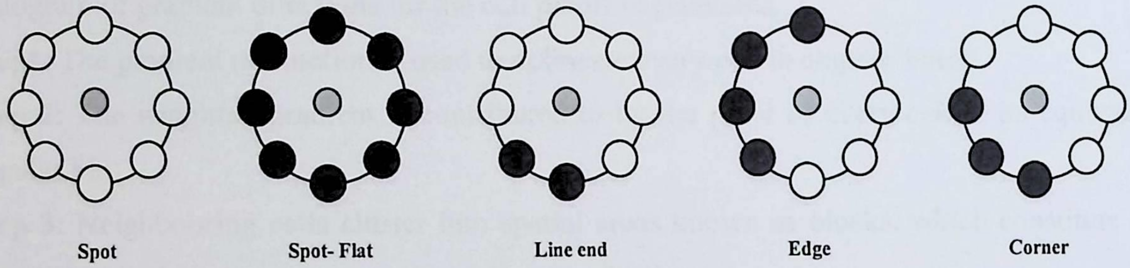



*Figure 2-11: Different texture primitives detected by LBP (Pietikainen, et al. 2011)*

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 and Guo 2015) (Pietikainen, et al. 2011).

**Histograms of Oriented Gradient (HOG)**

Objects in images can be identified with the feature descriptor called a histogram of oriented gradients (HOG). This method is designed to quantify gradient orientation occurrences in localised image sections. The underlying principle is that the arrangement of local intensity gradients or edge directions can be used to describe the shape and appearance of a local object, regardless of accurate data about equivalent gradient or edge positions. The practical application involves separating the image window into small spatial areas called cells, each having an aggregation of a local one-dimensional histogram of gradient directions or edge orientations over the cell pixels. The representation is the outcome of the totality of histogram entries. Furthermore, it is advisable that, prior to employing the local responses, they should be contrast-normalised to improve invariance to illumination, shadowing size ("blocks") and use of results to achieve normalisation of every block cell. The normalised descriptor blocks are considered to be HOG descriptors. The human detection chain is obtained by using a dense, actually overlapping, HOG descriptor grid for tiling the detection window as well as employing the combined feature vector (Dalal and Triggs 2005) (Patil, Junnarkar and Gore 2014).

As shown in Figure 2-12, the HOG descriptor extraction algorithm operates through a series of steps. First, the input image is separated into small interlinked areas (cells), and for each cell a histogram of gradient directions for the cell 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 block histogram is the outcome of a histogram group that has been normalised, and the descriptor is denoted by a block histogram set.

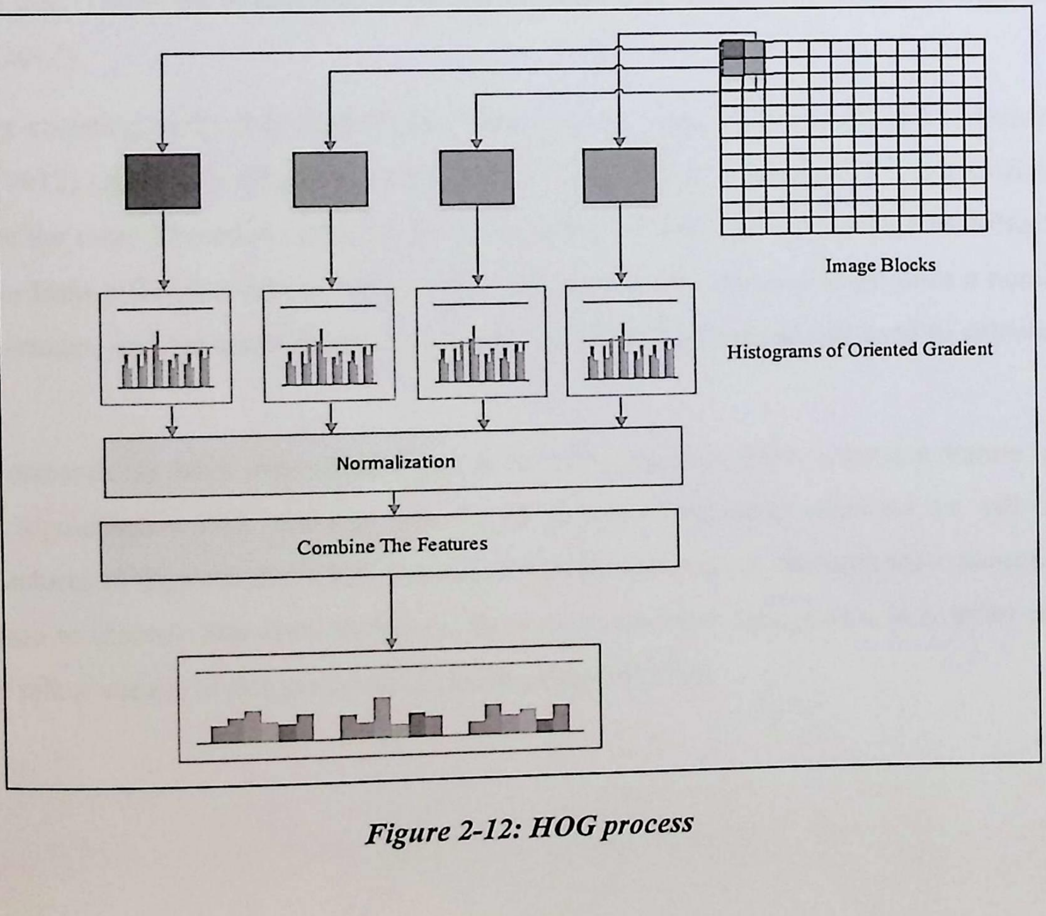



*Figure 2-12: HOG process*

**Multi-fractal Dimension Features (FD)**

To determine how complex an object shape or texture is, fractal dimension measurements can be performed in the context of the image analysis paradigm. Fractional dimensions are described by a fractal geometry based on a range of approaches, where Hausdorff's dimension is one such major fractional dimension. For an object with a Euclidean dimension E, the equation for calculating the Hausdorff's fractal dimension $D_0$ is:

$$D = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \epsilon^{-1}} \qquad (2\text{-}12)$$

In the above equation, the quantification of hyper-cubes of dimension D and length $\epsilon$ present within the object is denoted $N(\epsilon)$. The box-counting algorithm can be applied to estimate $D$ for $D_0$ if a binary image is taken as a representation of an object. The algorithm associated with the 2D case unfolds across several steps with no loss of generality. The first step is to partition the image into a grid comprising squares of size $\epsilon \times \epsilon$. Subsequently, the number $N(\epsilon)$ of squares of size $\epsilon \times \epsilon$ containing a minimum of one object pixel is determined. A $\log \breve{N}(\epsilon)$ vs. $\log \epsilon^{-1}$ curve can be formed through parametric variation of $\epsilon$. The last step is the use of a line-fitting technique, such as least squares fitting, to estimate the curve so formed as approximated to a straight line. The slope of this line is the fractal dimension FD (Costa, Humpire-Mamani and Traina 2012).

The box-counting method is used for the binary image, where (Costa, Humpire-Mamani and Traina 2012) (Alrawi, Sagheer and Ibrahim 2012) argue that a single FD is not sufficient to describe the case. Therefore, they proposed an effective method to produce a multifractal dimension from difference binary images. They created a new method to produce a number of binary images, and for each binary image the box-counting method has used to calculate the FD.

In this research we have used these features for two purposes. First, a texture feature is used for the segmentation task. For example, in our proposed trainable solutions we will use the HOG features to segment the image instead of the intensity values. Second, these features have been used to identify any abnormality in the case in question (pregnancy or ovarian case) to test the effectiveness of our proposed segmentation solutions.

### 2.3.4 Classification

Classification is extensively used in this research. It is a process of mapping data into predefined classes, which includes two stages. The first learning (or training) stage involves generating a classification model using training data. The second testing phase involves testing the performance of the classification model using test data. A classification model that has been implemented and tested properly with satisfactory level of accuracy can then be applied in practice the data objects with no class labels. In this work, we are particularly interested in two types of classifiers: support vector machine (SVM) and artificial neural network (ANN) because both are well established approaches that can handle feature vectors of high dimensionality.

**ANN Classifier**

An ANN is a network of interrelated computational units (nodes) as shown in Figure 2-13. Each node represents two processing functions: summarisation and transformation. The summarisation function calculates a weighted sum of its inputs where each input consists of an input value and its associated weight. A weight $w_{ji}$ normally refers to the multiplicative between the input node $j$ in the previous layer of the network and the output node I of the current layer. A transformation function, also known as an activation function, of the node applies a transformation, normally non-linear, to the result of the summarisation function.

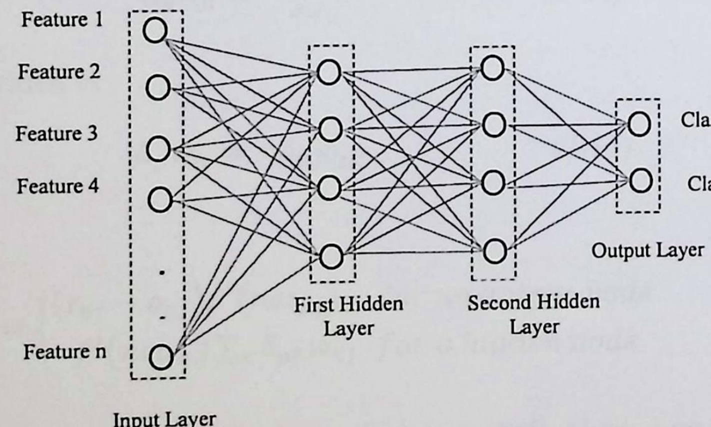



*Figure 2-13: A typical neural network topology. The features size presents the number of input nodes. The number of nodes and the hidden layers are determined experimentally. The number of output nodes detected based on the class number.*

The multilayer perceptron networks (MLPs) are neural networks used in pattern recognition problem. The MLP is a feed-forward neural network with multiple node layers in a directed graph. The MLP is recognized as among the ideal techniques for pattern recognition and in providing an adaptive learning ability. Furthermore, there are evidences that a universal approximator should have a two-layer backpropagation network with appropriate hidden nodes.

Back-propagation ANN's such as ones that have been used in this study involve an input layer, output layer and hidden layers (Zumray Dokur 2002) (Moghaddam and Soltanian-Zadeh 2011). The Training Phase is a mapping between the input and the output values. By using a learning algorithm (e.g., generalized delta rule), the mapping procedure is accomplished by adjusting the value of the weights. The error for the training phase is defined as follows:

$$E_p = \frac{1}{2} \sum_j (t_p - o_{pj})^2 \qquad \text{(2-13)}$$

In this equation, the index $p$ corresponds to single input vectore, the vectors $t_p$ is the target and $o_p$ observed output vectors according to the $p$. In general, the input vector is a vector of features values (ie LBP or HOG) and the target vector (binary class according to the target tissues), indicating the class number to which the pixel should be assigned. The error function has used to calculate the diffrence between the observed and the target vectors. The error can be minimized by changing the weight $w$ (Moghaddam and Soltanian-Zadeh 2011) (Egmont-Petersen, de Ridder and Handels 2002). To minimize that error, a gradient descent in $E$ has implemented by the generalized delta rule as follows:

$$\Delta_p w_{ji} \propto -\frac{\partial E_p}{\partial w_{ji}} \qquad \text{(2-14)}$$

which can be rewritten as

$$\Delta_p w_{ji} = \eta \delta_{pj} o_{pi} \qquad \text{(2-15)}$$

Where

$$\delta_{pj} = \begin{cases} (t_{pj} - o_{pj}) f'(net_{pj}) & \text{for an output node} \\ f'(net_{pj}) \sum_k \delta_{pk} w_{kj} & \text{for a hidden node} \end{cases} \qquad \text{(2-16)}$$

in which $\boldsymbol{net_{pj}} = \sum_k \boldsymbol{w_{ji} o_{pi}} + \boldsymbol{\theta_j}$, is the total input to node $j$ including a bias term $\theta_j$ and the parameter $\boldsymbol{\eta}$ is the learning rate. The output of node $j$ due to input $p$ is thus $o_{pj} = f_j(net_{pj})$ with $f$ the activation function. In the case of a sigmoid activation function $\boldsymbol{f(z)} = 1/(1 - e^{-x})$. Therefore the $\delta_{pj}$ can be rewritten as:

$$\delta_{pj} = \begin{cases} (t_{pj} - o_{pj})o_{pj}(1 - o_{pj}) & for\ an\ output\ node \\ o_{pj}(1 - o_{pj})\sum_k \delta_{pk}w_{kj} & for\ a\ hidden\ node \end{cases} \quad (2\text{-}17)$$

Finally, an additional momentum term can be added to the learning equation resulting in

$$\Delta_p w_{ji}[n] = \eta\delta_{pj}o_{pj} + \mu\Delta_p w_{ji} \quad (2\text{-}18)$$

where $\mu$ is the momentum rate At each iteration, the weights are thus modified as follow:

$$w_{ji}[n+1] = w_{ji}[n] + \Delta_p w_{ji}[n] \quad (2\text{-}19)$$

After the ANN is adjusted with specific weight values in the training phase, the trained ANN is used to classify unknown samples.

**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). SVM is a type of non-probabilistic binary classifier that have been used in many applications. The basis of this method is that decision plane, which categorises objects into a positive and negative class, defines decision boundaries. Figure 2-14 displays an example of the decision plane, where a boundary is shown between the red colour and black 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 black. The line that separates the two classes is referred to as a hyperplane. Two classes can be separated by many different types of the hyperplane. 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 2-14, the optimal separating hyperplane is shown as a blue line. The features along the boundaries are called "support vectors". It is the support vectors that are used to differentiate the two classes. Figure 2-14 presents an overview of the SVM process.

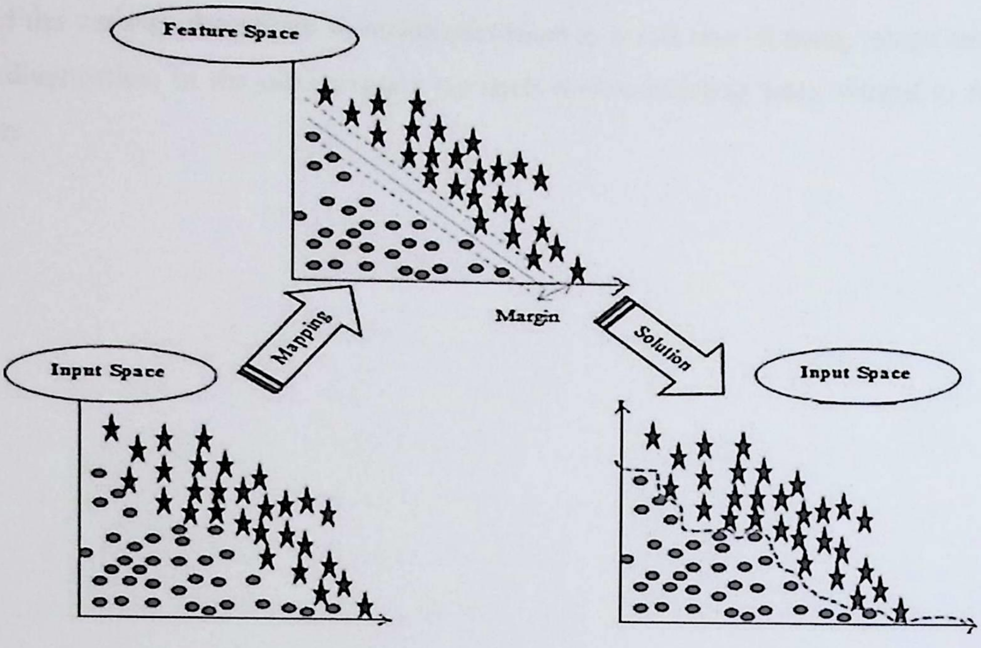



*Figure 2-14: Binary SVM classifier process (Tomar and Agarwal 2015)*

The algorithm to classify a new case start by calculate the score (Sc) of the test case $x$ is based on the Lagrangian formulation:

$$Sc(x) = \sum_{i=1}^{n} \alpha_i k(Si, X) + b \qquad (2\text{-}20)$$

Where $Si$ is a support vector, $\alpha_i$ is the weight of Si, $n$ is the number of support vectors, and $k$ 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. If $Sc \geq 0$, then $x$ is classified as positive, otherwise it is classified as negative.

We close this section by pointing out that ANN is a more effective classifier when we have sufficient samples to train. This is also true for any classifier, but SVM is suitable when dealing with high dimensional feature vectors. In this thesis, we use ANN when we have enough training samples and SVM otherwise.

## 2.4 Summary

The background to the various medical and technical issues relevant to the research carried out in the thesis was given in this chapter to equip the reader with basic understanding of the tools used in general medical image-based diagnosis. The focus was on the information on the use of US imaging system for gynaecological diagnosis abnormalities, the specific technical challenges that will be met in designing effective computer-based solutions for both miscarriage and tumour classification. The challenges described were limited to image pre-processing (noise reduction), segmentation methods and the extraction of relevant texture feature. We also described the various classification techniques used as a last task of using image analysis for medical diagnostics. In the next chapter we shall review existing work related to this thesis objectives.

# Chapter 3. **Literature Review**

Having discussed and reviewed the various concepts and image processing/analysis tools that have been developed for general biomedical imaging application, we now turn our focus on the existing research conducted in the analysis of US ovary scan in search of gynaecological abnormalities. In this chapter, we therefore review and describe recent influential work about our research area of interest. The lesson learnt from last chapter is our motivating guide for this review to focus on the following main technical tasks: segmentation of ultrasound images for miscarriage cases and the detection of ovarian tumour classification. We shall also review the literature on performance measure that are relevant in evaluating the effectiveness of any algorithms to deal with the challenges relevant to our work.

## 3.1 Ultrasound Image Segmentation for Early Pregnancy Monitoring

There are a number of approaches that have been proposed to segment the ROI in US images for detecting different types of diseases (Noble and Boukerroui 2006) (Meiburger, Acharya and Molinari 2017). The research has, however, been very limited in the area of segmentation for gestational sacs from US images. This section will mainly discuss semi-automated or automated segmentation algorithms for extracting and measuring the GS and foetal. Table 3-1 summarise the existing methods that have been used to segment the GS and foetal objects.

***Table 3-1: Summarise the existing approaches for both GS and foetal segmentation.***

| Reference | Title | Technique | NO. Images | Result |
|---|---|---|---|---|
| **(Khazendar, Al-Assam, et al. 2014)** | Automatic Identification of Early Miscarriage Based on Multiple Features Extracted from Ultrasound Images | semi-automatic segmentation method based on the Otsu threshold. | 68 | There is no evaluation for the segmentation stage. |
| **(Khazendar, Farren, et al. 2015)** | Automatic Identification of Miscarriage Cases Supported by Decision Strength Using Ultrasound Images of the Gestational Sac | Histogram equalization followed by the Otsu threshold method.<br>Semi-automatic approach | 184 | $R^2$ between the manual and automatic solution =0.98. |
| **(David H. Hareva 2016)** | Automatic Gestational Age Estimation Based on Crown Rump Length and Gestational Sac. | Gaussian kernel followed by the Otsu threshold. | 10 | CRL accuracy = 60% and MSD accuracy= 70%. |

| | | | | |
|---|---|---|---|---|
| **(Supriyanti, et al. 2018)** | Measuring gestational age and uterine diameter based on image segmentation. | set the threshold manually, morphology operations and edge detection | Have not mentioned | Error value = 5.3%. |
| **(Zhanga, et al. 2011)** | Automatic measurement of early gestational sac diameters from one scan session | Viola-Jones method | 31 videos | An error of 0.059 and 0.083 for depth and length respectively |
| **(Zhang, et al. 2012)** | Intelligent scanning: Automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination | intelligent scanning | 31 videos | Depth and length (diameter) measurements were 7.5% ±5.0%, 6.5%± 4.6%, and 5.5%± 5.2%, respectively. |
| **(Imaduddin, et al. 2015)** | Automatic detection and measurement of fetal biometrics to determine the gestational age | Adaboost-Randomize Hough Transform | 300 biparietal head images and 200 femur images | A quite low detection rate of 44 out of 300 images for the foetal head. A detection error rate of 18 out of 200 of the foetal femur. A Good correlation between the manual and automatic solution. |
| **(Ni, et al. 2013)** | Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters | AdaBoost model and iterative randomized Hough transform | 675 images. Training= 500 Testing =175 | The mean-signed difference between the manual and the automatic solution= 1.6% |
| **(Gustavo, et al. 2007)** | Automatic fetal measurements in ultrasound Using constrained probabilistic boosting Tree | Boosting tree classifier | Training =1, 426 samples. Testing = 177 ultrasound images of fetal head, 183 of fetal abdomen, and 171 of fetal femur | Average error of 0.0265 between the manual and automatic solution. |
| **(Rawat, et al. 2013)** | Automatic assessment of foetal biometric parameter using GVF snakes | Gradient Vector Force (GVF) | 12 images | Showed the result as a segmented image. |
| **(Dahdouh, et al. 2015)** | Segmentation of embryonic and fetal 3D ultrasound images based on pixel intensity distributions and shape priors | A shape constrained multiphase level set segmentation method | 14 3D images | The average level of similarity between the manual and automatic solution, sensitivity and specificity are: 0.8, 0.7, 0.75 |

| | | | | |
|---|---|---|---|---|
| | | | | and 0.98, respectively. |
| (Lu, Tan and Floyd 2005) | Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform | K-means clustering and iterative randomized Hough transform. | 217 images | The ROI were detected in 214 out of 217 Images. Differences between the manual and prosed solution were 0.12 for BPD and -0.52 for HC. |
| (Yu, et al. 2008) | Fetal abdominal contour extraction and measurement in ultrasound images. | fuzzy means clustering and iterative randomized Hough transform | 150 images | Accuracy= 98% . Mean absolute difference= 1.2% The standard deviation =2.1%. |

### 3.1.1 **Thresholding**

The effectiveness of the basic thresholding principle is limited for ultrasound images of this problem domain due to the complexity of the associated images. Therefore, several authors have combined thresholding with other techniques, such as pre-processing and post-processing, to avoid a number of the complexities relating to speckle noise and to deal with false positive ROIs. Based on our research, most of the existing works are semi-automatic or manual in nature. For example, (Khazendar, Al-Assam, et al. 2014) and (Khazendar, Farren, et al. 2015) proposed an algorithm to measure the diameters of the GS and use the measurement results and some other geometry features (such as volume) of the sac to diagnosis miscarriage cases. The algorithm relies on manually cropped ROI (i.e., GS) from the start. Although the MSD, volume and perimeter features are measured accurately with satisfactory classification results (an overall accuracy of 98% with sensitivity (miscarriage) 97% and specificity (normal) 99% was reported), the very fact that the GS has to be manually segmented highlights the need for an automatic segmentation solution rather than solutions for segmentation.

Recent work using the threshold was presented in (David H. Hareva 2016) and (Supriyanti, et al. 2018). The reported method obtains the Crown-Rump Length (CRL) and MSD values automatically to determine the age of the foetus. The method uses the Gaussian kernel and a hump-shaped Gaussian to enhance the image and reduce the effect of speckle noise. Morphological operators are then used to further consolidate the ROI, and the Otsu thresholding is used to eventually extract it. On a primary empirical study using only 10 sample ultrasound images, the successive calculation of CRL and MSD parameters on the segmented ROIs was, respectively, 60% and 70%. Both the levels of accuracy and the scale of the study were extremely limited, and the results were only indicative than conclusive. More recently, (Supriyanti, et al. 2018) proposed a method to detect the ROI using template matching. This method takes a trample from the input image, divides the image into six blocks, and then crops the best window that can cover the whole object. This process is followed by applying a manually set threshold, morphology operations and edge detection to eventually determine the ROI. The age of the pregnancy is estimated using the measured MSD after the ROI cropping. The estimated gestational age has an error value of up to 5.3%.

### 3.1.2 Machine Learning-based Segmentation

This approach to segmentation relies on the features describing the microstructure of the tissue being imaged and is consequently classifying the regions or the pixels of the image into a different type of classes based on the texture features. The idea of using texture features is related to the fact that textures depend on statistical patterns of intensity, and hence more independent of the imaging system physics. An example of using this approach for automatic segmentation can be found in (Zhanga, et al. 2011) where a solution was proposed to measure the diameter of the early gestational sac from a 2D ultrasound image video. The solution consists of four principal stages: training, detection, indexing and measurement. The Viola-Jones method has been used in the training stage. This method involves two Cascade AdaBoost classifiers. The two trained classifiers detect candidate GSs at the frame level from a sliding window over a sequence of a US frame created by any scan session video. Given that detection outcomes mostly have false positive (Non-GS), an indexing method which utilises local context information about the relative position between anatomies in the image sequence to first select the frame containing the maximal candidate GSs as the standard plane and then eliminates the false positive GSs. An error of 0.059 and 0.083 for depth and length respectively was recorded on experiments conducted on 31 videos. This work could arguably be considered the first attempt at the fully automatic segmentation of GS, showing the promise of a machine learning-based trainable segmentation approach.

This work has been developed to produce the first completely automatic solution for selecting a standardized plane for an early gestational sac (SPGS) and performing biometric measurements using real-time 2D ultrasound scan data (Zhang, et al. 2012) where an automatic method entitled “intelligent scanning” was proposed. The method included a sequence of steps, where, first, the Gestational sac is accurately and efficiently located in each ultrasound frame by developing a course to fine detection framework grounded on the training of two cascade AdaBoost classifiers. Second, the Non-SPGS (false positive cases) are automatically removed. Lastly, a database-guided multiscale normalized cuts algorithm is implemented to produce an initial segmented image (initial contour). This solution was tested on 31 ultrasound videos which were retrieved from the results of 31 anonymous patients. The differences between the manual measurement and the automatic solution with respect to SPGS selection, depth and length (diameter) measurements were 7.5% ± 5.0%, 6.5% ± 4.6%, and 5.5% ± 5.2%, respectively.

Another focus in pregnancy US scan is in the automatic trainable segmentation and measurement of foetal biometrics presented in (Imaduddin, et al. 2015). A method to automatically segment and measure the head circumference (HC) and automatic measurement of femur length was proposed. The method consists of pre-processing, feature extraction and classification stages. The cropping and scaling phases have used as pre-processing to reduce background noise i.e. the false positive objects from the input ultrasound images. In the feature extraction stage, a Haar-like four type basis function is used to get the optimal configuration to be used in the training phase. Lastly, at the classification stage, an Adaboost-Randomize Hough Transform (RHT) method was proposed to segment and measure the foetal head, and an Adaboost-morphology scheme was employed to detect and measure femur length. The proposed methods were tested using 300 biparietal head images and 200 femur images. The study showed a quite low detection rate of 44 out of 300 images for the foetal head, and good measurements of biparietal HC with a correlation coefficient of 98% and an average error of 0.039 between the manual measurement ground truth and the automatic measurements. With regards to the detection and measurement of the foetal femur, a detection error rate of 18 out of 200 ultrasound images with an average error of 0.101 and a correlation coefficient of 0.763 between the manual and the automatic measurements was reported. The work actually demonstrates the number of difficulties inherent to detecting the correct ROI in ultrasound images.

(Ni, et al. 2013) also reported a machine-learning based method to segment the foetal ultrasound image followed by detecting the head and measure the circumference of the head of the foetal. The AdaBoost model based on the Haar-like features has been used, and then, for the first time, the prior knowledge and online imaging parameters have been used to control the sliding window-based head detection from US images. Their method can significantly enhance both speed and detection accuracy. A local phase-based method, which is invasive in the pixel intensity changes and the speckle noise in the US images, is used to detect the edge of the head in the localised image. Lastly, to detect the ellipse of the head contour, an iterative randomized Hough transform (IRHT) has been used. Experimental results on 675 ultrasound images (500 images for the training phase and 175 for testing phase) demonstrate the effectiveness of the proposal in producing accurate measurements. Where the mean-signed difference between the manual and the automatic solution measurements is 2.86 mm (1.6%).

(Gustavo, et al. 2007) proposed a new automated solution that measures the femur lengths (FL) of foetal femurs from US images and the HC and biparieter (BPD) of the foetal head from the US images. The proposed solution uses a discriminative constrained probabilistic boosting tree classifier to identify the ROI from the background. This model has the ability to filter out un-wanted area. The experimental results confirm that the proposed framework yields automatic measurements for the FL and the BPD very close to a gynaecologist's manual measurements with average error 0.026, without inter- and intra-observer variabilities.

### 3.1.3 **Deformable Model-Based**

The deformable models have been widely used in the field of ultrasound image segmentation for different applications. Deformable model on the definition of internal and external energy and the evolution of an initial contour until the two energy functions reach a balance. The deformable model has been used in (Rawat, et al. 2013) for a reliable semi-automatic segmentation scheme to extract and measure the dimension of the GS. The power of this model lies in employing the Gradient Vector Force (GVF) to deform the initial contour (initial mask) and grow it to fit the actual border of the ROI. In this work, the proposed model was applied to 12 images, and the authors showed the result as a segmented image to illustrate the accurate segmentation of the proposed method.

### 3.1.4 **Clustering**

Clustering is a mathematical technique that aims to partition sets of objects (e.g. image pixels) into groups sharing some similarity properties called clusters. Given an image, we can use a clustering algorithm to classify each pixel into a specific group depending on its relevant to our objective. In theory, image pixels that are in the same cluster should have similar features and/or properties, while images pixels in different clusters should have highly dissimilar features and/or properties. Clustering is an unsupervised learning technique and is one of the most popular methods used for statistical data analysis in many fields (Clarke 1996). The clustering method was used in producing the initial segmentation of RoI. In (Dahdouh, et al. 2015), a shape constrained multi-phase level set segmentation method is described to segment the 3D ultrasound images and extract the foetal envelope. The initialization of the process is semi-automatic segmentation was performed by asking the expert to detect the CRL and using it to identify the location of the ROI. The k- means clustering method was used to produce the initial segments. Based on the location and the initial segmentation, the whole ROI is extracted and tested on a database of 14 3D US. The experiment results have demonstrated that proposed method has its promise in assisting accurate segmentation compared to the manual segmentation. They have used Dice method to evaluate the closeness between the manual and expert measurement. The average levels of Dice, similarity, sensitivity and specificity are: 0.8, 0.7, 0.75 and 0.98, respectively.

Another example of using the K-means clustering technique can be found in a work conducted by (Lu, Tan and Floyd 2005). The objective of this research was to design an automatic solution to extract and measure the BPD and HC from the US foetal image. New object detection model has proposed. A K-means clustering method has been used as an initial stage to cluster each pixel according to intensity value (grey level) to produce a binary image. Because the foetal skulls are not completely closed, foetal head contours usually contain moderate to large gaps. An iterative randomized Hough transform (IRHT) was designed to estimate curve for the large gap. Automatic measurements were considered in relation to manual measurements to assess the degree to which the proposed system was effective and the differences between them were 0.12 for BPD and -0.52 for HC.

(Yu, et al. 2008) fuzzy means clustering and IRHT were used to design a method that can help to segment and measure the foetal abdominal contour from US images. Foetal abdominal circumference (AC) is an important measurement that can help to identify the abnormality during antepartum ultrasound monitoring. The fuzzy means as a clustering method is employed to distinguish clear borders, due to the contour of the abdominal tissues, from poor borders, attributable to the other region. Subsequently, to detect the outer border of the abdominal contour and produce an initial segmentation and estimation of the AC, the IRHT has used. Finally, the GVF method is used to fit an ellipse to the real border of the contour of the abnormal borders. The proposed method has achieved around 98% as accuracy of the segmentation.

## 3.2 Ultrasound Image Segmentation for Ovarian Tumour Diagnosis

In ultrasound imaging, ovarian follicles are hypoechoic structures found in the ovary whereas an ovary is generally obtainable as a medium amount and an identical structure. According to doctors, it is recommended to know the numeral value of follicles and compute their diameters.

Follicle walls should also be analysed in the clinics. Sometimes follicles can overlap within an ultrasound image, thus increasing counting mistakes and making the automatic segmentation method were considered to be difficult. Furthermore, differences in follicle shape and the presence of artefacts, particularly if the hyperechoic formation is there on top of the image, add more challenges to automatic segmentation process. More articles published in the literature were addressing these challenges. The existing approaches that have been used to segment the ovarian ROI have been summarized in Table 3-2.

***Table 3-2: Summarise the existing approaches for ovarian mass segmentation***

| Reference | Title | Technique | NO. Images | Result |
|---|---|---|---|---|
| (Padmapriya and Kesavamurthy 2016) | Detection of follicles in poly cystic ovarian syndrome in ultrasound images using morphological operations. | Enhance the threshold by using entropy optimizer followed by the morphology operation. | 12 images | The average recognition= 87.5%. |
| (Saranya and Maheswari 2016) | Follicle detection in ovary image using adaptive particle swarm optimization. | adaptive particle swarm optimization and Otsu method. | 158 images | Recognizes all follicles correctly. |
| **(Potocnik and Zazula 2000)** | Automated ovarian follicle segmentation using region growing | Region growing | 50 images | Recognition rate = 78%. |
| **(Krivanek and Sonka 1998)** | Ovarian ultrasound image analysis: Follicle segmentation | Watershed Transform | 36 images | $R^2 = 0{:}85 \pm 0{:}96$. |
| **(Zhang, et al. 2016)** | A region-based segmentation method for ultrasound images in HIFU therapy | Active Contours | 50 images | The average of the similarity, TP and the FP were 90.21% , 94.42% and 4.71%, respectively. |

## 3.2.1 Thresholding

In this problem domain, due to the complexity of the images, the best threshold is often found by using an optimization technique such as entropy (Padmapriya and Kesavamurthy 2016) and swarm optimization (Saranya and Maheswari 2016). In (Padmapriya and Kesavamurthy 2016), an accurate method of detecting PCOS (Poly Cystic Ovarian Syndrome) is described, which involves segmentation of follicles from ultrasound images, which included operations such as image histogram equalization, filtering, thresholding (using the entropy optimizer method), area opening, closing and merging. Then, the number of follicles is counted and classified as

PCOS present/PCOS absent. The method was tested on 12 images. The average recognition rate for all images processed by the proposed method was 87.5%.

In (Saranya and Maheswari 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) methods. The problem of thresholding is reduced to an optimization problem 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.

### 3.2.2 Region growing

Region growing method was employed for the ovary mass segmentation problem (Potocnik and Zazula 2000). The paper proposed a fully automated recognition system that is mainly based on three steps. The first step involves the determination of the initial homogenous regions using two independent procedures. In the first procedure, the initial or original image is smoothened by an adaptive neighbourhood median filter. The second procedure involves the calculation of the standard deviation of each pixel based on its neighbourhoods. If the standard deviation of a pixel does not exceed the threshold H and its grey-level is smaller than the image's grey-level, then the image is marked as homogenous. The average grey-level controls the growing through the weighted image gradient that has been newly introduced. Regions that correspond to the follicles are then extracted. A total of 50 ovarian US images have been used to test this algorithm. It was found out that the recognition rate was at 78% when this procedure is used.

The two methods, region growing and thresholding, were found to perform well but they still need segmentation refinement. The refinement was necessary since thresholding an image will always include unwanted objects of the image. Morphological operations can, however, be applied to mitigate this issue or through extraction of the required features from the segmented objects. The majority of articles found in the literature focussed purely on the follicle inner border segmentation.

### 3.2.3 Watershed Transform

This section focuses on a method proposed by (Krivanek and Sonka 1998) about 20 years ago. This method is called Watershed transform and is able to segment both the outer and inner borders of the ovarian follicles. It does this through an approximation of the inner borders of the ovarian follicles in the study. The approximated inner borders are used to automate the process of defining the region of interest (ROI). This method was tested using 36 US images of ovaries. The validation has shown a great relationship between the manually defined area measurements and the computer-determined ones. While the research presented herein delivers interesting results and it is a technique that could now be rendered completely automatically, many advances in finding the inner and outer boundaries of a follicle were found in the recent literature. Considering another problem domain, that of uterine fibroid segmentation, a considerable amount of literature on accurate fibroid segmentation has been found that describes how to correctly guide High Intensity Focused Ultrasound (HIFU) treatment. US images show Fibroids as hypoechoic regions. This poses the problem of dealing with the fact that HIFU guidance produces low-quality images since they have high signal-to-noise ratio as compared with US images.

### 3.2.4 Active Contours

Active contours methods were used extensively in the segmentation process, (Zheng, Liu and Zhu 2013), (Ni and Yuan 2015) (Ni, et al. 2016). This method, however, requires that the target area is inserted manually or through the use of sequences of images based on the US image. (Zhang, et al. 2016) presented a new recent work that is automated completely. The method is different from those discussed in the literature since it is based on the division of an image into its superpixels and homogeneous regions which are used to extract the target features. The method was tested using 50 US images and showed incredible results with the average of the similarity, TP and the FP were 90.21% , 94.42% and 4.71%, respectively.

## 3.3 Performance Evaluation methods

As the problems addressed in this thesis are of different in nature, a number of metrics are used to evaluate the performance of proposed solutions. In this section, the evaluation metrics and the methods by which they are applied are described. A general understanding of the metrics is needed before empirical evaluation of the performance of the proposed solutions is undertaken.

### 3.3.1 Using Statistical Methods for Measurement Closeness

In the problem domain of detecting miscarriage in early pregnancy, the effectiveness of ROI segmentation can be tested by the closeness of the automatic measurements of the ROI to manual measurements on the one hand, and the correlation between the manual and the automatic measurements on the other hand.

**Linear Regression for Correlation**

Fitting a regression line helps to characterise how actual and predicted measurements are correlated. The correlation between these two measurements can be assessed with the statistical measure known as coefficient of determination $R^2$, which thus indicates the proximity of the data to the fitted regression line (The Minitab Blog 2016). The formula for determining $R^2$ is:

$$R^2 = \left(\frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}}\right)^2 \qquad (3\text{-}1)$$

In the formula, $x$ and $y$ denote two variables, while the number of observations is denoted by $n$. The actual and predicted measurements are better correlated the higher the $R^2$ value is. Although it elucidates how two variables are related, $R^2$ is not informative about whether or not the predicted variable has bias. This issue can be dealt with by employing the Angle of Regression Line (ARL), which sheds light on the character of the correlation between two variables. In general, automatic and manual measurement values are correlated well if the ARL is about 45°. By contrast, automatic measurement values have bias if the ARL is below or above 45°. Nevertheless, closeness is not indicated by either correlation or ARL. Despite being indicative of the manner of correlation between two types of measurement, $R^2$ cannot offer information about the closeness of those measurements or the level of consistency of their values (Bland and Altman 1986).

**The Bland and Altman Method for Closeness**

This is a scatter plot method invented by Bland and Altman (B & A) that describes the agreement or disagreement between two measurements measured quantitatively. This method, B&A is used to calculate the amount of agreement by constructing limits of agreement between measurements. The differences in the mean and standard deviation of two quantitative measurements are the two major methods of calculating the statistical limits. The scatter plot has two axes where the Y-axis are used to illustrate the difference between the two paired measurements (A-B) and the X-axis represents the average of these measures ((A+B)/2). That is to say, the difference of the paired measurements is plotted against the mean of the two quantitative measurements. The B&A method considers that two measurements are indeed close if 95% of the data points should lie within $\pm 2\sigma$ of the mean difference (Bland and Altman 1986). This is based on the normal distribution theory.

### 3.3.2 Measures for Classification Accuracy

In the problem domain of segmenting ovarian tumour images for counting the number of follicles, we test the effectiveness of the segmentation through running a complete CAD cycle. In other words, we want to extract texture features from the segmented ROI, build a classification model using training examples of the extracted features, and then run tests on test examples and measure the level of accuracy of the predictive classification model. The classification model is a classifier that maps the instances from the outcome results to the expected classes. The classification models can either produce continuous output while others predict discrete classes of output. The labels {Y, N} are used to distinguish between the actual and the predicted class. There are four possible outcomes given a classifier:

1- True Positive TP *(correctly identified)*: Sick cases correctly diagnosed
2- False Positive FP *(incorrectly identified)*: Healthy cases incorrectly identified
3- True Negative TN *(correctly rejected)*: Healthy cases correctly identified
4- False Negative FN *(incorrectly rejected)*: Sick cases incorrectly identified

With a set of testing instances and a classifier, a contingency table which is in the form of a two-by-two confusion matrix can be constructed to represent dispositions of the instances of a set.

| | | Actual | |
|---|---|---|---|
| | | P | N |
| Predication | Y | **True Positive** | **False Positive** |
| | N | **False Negative** | **True Negative** |

$$Sensitivity = \frac{TP}{TP+FN} \tag{3-2}$$

$$Specificity = \frac{TN}{FP+TN} \tag{3-3}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3\text{-}4)$$

## 3.4 Summary

Here we completed the review of the background material and existing work that were necessary for our endeavour to design and evaluate solutions to the various tasks to be carried out in this research product. In the remaining chapters of this thesis we shall follow the roadmap identified here which consist of designing efficient and highly accurate segmentation of the region of interest, extracting appropriate texture features to support the segmentation and to be input into appropriate machine learning diagnostics models, and evaluating the developed schemes.

# Chapter 4. Conventional approaches to Segmentation of Gestational Sac

This chapter is mainly concerned with automatic segmentation and measurement of the gestational sac in relation to the classification of cases of miscarriage in early pregnancy. The content comes into two parts: the first part seeks to, explore the full potential of threshold-based segmentation techniques due to their simplicity and ease of implementation; whereas the second one aims at, appreciating the problems faced by threshold-based methods and the degrees of difficulty involved even for this relatively simple application scenario where the ROI (GS) is much less complicated than ovarian tumour images, and hence identifying the fundamental limitations of the threshold-based solutions, paving the way for the introduction of more advanced machine learning-based techniques in segmenting the ROI in more complex ultrasound images.

Based on the investigation and the review of the current literature, it has been found that speckle noise is one of the critical factors that can affect the accuracy of segmentation and diagnosis. Therefore, in this chapter, it is significant to review the more effective filters for suppressing the speckle noise and enhancing ultrasound images for more effective segmentation of the ROI. Based on the review, we will also focus on use of wavelet transform as a speckle noise filter to supress noise and highlight the object of interest, which ultimately makes threshold-based segmentation more effective. In addition, there is a focus on the Wiener filter that can help to reduce speckle noise without too much damage to the important texture information for machine-based diagnosis. Then evaluating some traditional threshold-based segmentation techniques and highlight factors that influence their performances and shortcomings will be conducted. Based on the evaluation, we further propose a new algorithm to automatically segment a GS from a static B-mode image by exploiting its geometric characteristics. the effectiveness of the proposed solution will be evaluated by firstly comparing the automatic size measurements of the segmented GS against the manual measurements, and then integrating the proposed segmentation solution into a classification framework for identifying cases of miscarriage and pregnancy of unknown viability (PUV). Test results demonstrate the merits of the proposed method in segmenting the GS with distinct geometric characteristics with high levels of accuracy on the one hand, and the limitations of the proposed algorithm in segmenting images with complex and confusing geometric characteristics on the other.

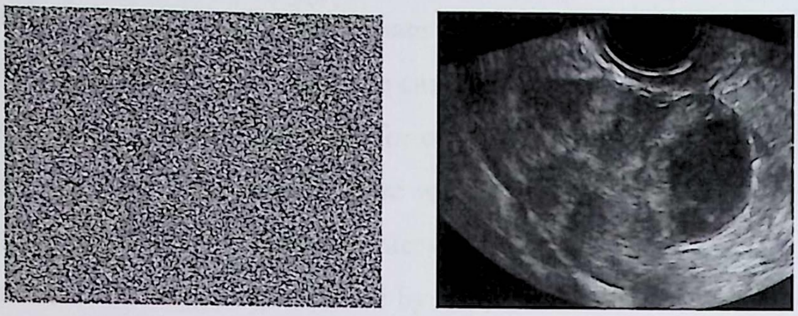
The chapter is organised as follows. The problem statement and the challenges to enhancement and segmentation of ultrasound images are first described in Section 4.1. Section 4.2, will evaluate the known speckle noise filters and focus on filters that are suitable for the segmentation of ROI and for texture feature enhancement. Section 4.3, is concerned with evaluating a number of existing segmentation methods by applying them to the ultrasound images of GSs. Section 4.4 then presents a threshold-based algorithm for automatic segmentation and measurements of GSs. The algorithm is later empirically evaluated using a dataset formed from miscarriage and PUV cases in Section 4.5. The difficulties faced by threshold-based solutions will be discussed in Section 4.6 before summarising the findings of the work presented in this chapter.

## 4.1 Problem Statements

This chapter aims to tackle two problems: (a) suppressing speckle noise and (b) automatic segmentation of GSs of various sizes. Both problems involve processing B-mode ultrasound images.

### 4.1.1 Speckle Noise Reduction

As we explained in section 2.3.1, ultrasound images are known to be corrupted by a special type of noise known as *speckle noise*, as caused by interference from randomly distributed scattering (Zhu, et al. 2009). Figure 4-1 shows the speckle noise and its effect on ultrasound images of an ovarian tumour. Speckle noise often has a negative impact on image quality by, in particular, hiding important details and blurring edges and, therefore, affecting the image segmentation and other post-processing operations, and may eventually reduce the overall diagnostic value of the image (Loizou and Pattichis 2008) (Zhu, et al. 2009).
Speckle noise reduction is a particularly important requirement for automatic processing and analysis of ultrasound images. However, this type of noise belongs to the multiplicative noise type. Applying the wrong type of filter, for example those filters suitable for additive noise reduction, may remove important information that can be helpful to in the identification of abnormal cases.



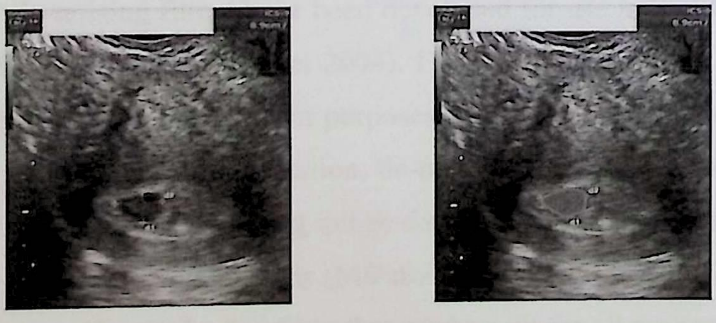
A) B)

*Figure 4-1: Illustration of speckle noise and its effect on an ultrasound image of an ovarian tumour: A) Example of Speckle noise (Dangeti 2003) and B) Ultrasound image of an ovarian tumour corrupted by speckle noise.*

In the context of this thesis, noise removal and image enhancement serve the objective of successfully segmenting GS. Therefore, any noise reduction and image enhancement methods must help to highlight the ROI (GS), remove irrelevant artefacts and eliminate false positive objects. In this case, a trade-off may have to be made to gain a good outline of the ROI borders, even at the expense of losing texture information within the ROI. However, in order to extract useful texture information after the segmentation step to allow for effective diagnosis of the abnormality (see chapter 7), such a trade-off cannot be easily made; too much useful information may be lost, and the accuracy of the diagnosis handicapped.

### 4.1.2 GS Segmentation

The automatic segmentation of GS refers to the automated process of extracting the GS region from the image with high precision. Precision, in this context, means that the segmented ROI should not include irrelevant parts outside, but at the same time should not have any significant parts of the ROI missing from, the segmented region. Figure 4-2 shows an example of a successfully, and precisely, segmented GS. The segmented region is marked by red pixels.



A) Original Image B) Segmented ROI: GS

*Figure 4-2: Segmentation of the GS and the ovarian tumour mass: A) the original GS image and B) segmented image*

The segmented GS region can be used to estimate the discriminating MSD parameter, the value of which is a critical diagnostic of miscarriage cases (as explained in chapter 2). Currently, the detection and measurement of the diameters for calculating the MSD for the GS are measured manually by ultrasound image observers. Here we investigate and develop techniques for the automatic segmentation of GS to overcome inter- and intra-observer variation. Unfortunately, successful segmentation is adversely affected by a number of factors including quality of the image, poor and blurred GS border, and similarity in the textures of the ROI and its background. Figure 4-3 illustrates these problems. In Figure 4-3-A, due to the presence of noise, the border region of the ROI is extremely blurred. The blurry effect creates a technical challenge for segmentation methods. Figure 4-3-B shows an area of the image where the boundary of the ROI is not clearly identifiable, i.e., the pixel values within the boundary and the pixel values outside the boundary are similar. In Figure 4-3-C, the texture inside the ROI is similar to the texture outside the ROI, which also causes difficulty in separating the ROI from its background.

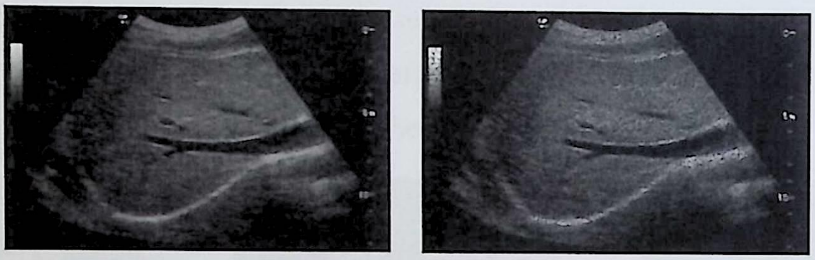
A) image quality B) Poor border C) Similarity between the texture

***Figure 4-3: Examples of challenges for correct segmentation of the ROI: A) image quality, B) poor border and C) similarity between the texture of the ROI and the background***

## 4.2 Speckle Noise Filters Evaluation

Image de-noising is a known problem in image processing that is usually resolved by applying noise filters. Many de-noising filters have been developed for use in both the spatial and frequency domains (Buades, Coll and Morel 2004). From the current literature, it is clear that different noise filters are used for different purposes such as de-noising the image in order to highlight the ROI for automatic segmentation, de-noising the image to highlight anatomical structure inside the ROI and/or facilitating image details for manual diagnosis, and enhancing the image textures for automatic diagnosis (Malathi and Shanthi 2010) (Thaipanich and Kuo 2010). There is no a specific method or filter that works well for all purposes and for all types of ultrasound image because images can be captured by different sonographers with different levels of experience on different machines with different setups. In this research, we will focus on the effective filters that can help ROI segmentation and enhance texture features for diagnosis.

In this section, we will first compare four well-reported noise filters: (a) median filter, (b) Fourier Butterworth filter, (c) wavelet transform, and (d) Wiener filter. To evaluate the effectiveness of the filters, we will use an image with only a small amount of noise in Figure 4-4-A as the reference image, and the noisy image in Figure 4-4-B which was generated using the MATLAB *imnoise* function. The function adds multiplicative noise using the equation $J = I + n * I$ where $n$ is uniformly distributed random noise with a 0 mean and 0.05 variance. The multiplicative noise added an effect that is essentially identical to the effects seen with speckle noise.



A) Reference Image B) Noisy Image

***Figure 4-4: A) Reference image and B) noisy image for evaluating noise filters***

The filters are evaluated by visually inspecting the output images with the filter applied and by using metrics including mean-square error (MSE), peak signal-to-noise ratio (PSNR) and signal-to-noise ratio (SNR) to measure image quality when the output images are compared against the reference image (Mateo and Fernandez-Caballero 2009).

$$MSE = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [I(m,n) - \hat{I}(m,n)]^2 \qquad (4\text{-}1)$$

$$SNR = 10 \times log_{10} \frac{\frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{n-1} I^2(m,n)}{MSE} \qquad (4\text{-}2)$$

$$PSNR = 10 \times log_{10} \frac{255^2}{MSE} \qquad (4\text{-}3)$$

In these expressions, $I$ is the original image and $\hat{I}$ is the estimation of the original image obtained from a noisy image. The images' measurements are $M \times N$.

The image quality metrics in Table 4-1 show that image quality degrades as window size increases, suggesting that applying the median filter to a small window size will give a better-quality image without losing too much important information.

*Table 4-1: Shows the MSE, PSNR and SNR for the Median Filter*

| Window | MSE | PSNR | SNR |
|---|---|---|---|
| 3 | 0.000772 | 79.25664 | 62.90485 |
| 5 | 0.001077 | 77.80678 | 61.45499 |
| 7 | 0.001434 | 76.56391 | 60.21211 |
| 9 | 0.001797 | 75.5851 | 59.23331 |
| 11 | 0.002204 | 74.69945 | 58.34766 |

Figure 4-5 presents the images resulting from application of the median filter with different window sizes. A closer visual inspection of the output images reveals that when the median filter is used over window sizes of 3×3 and 5×5, respectively, speckle noise is suppressed whilst at the same time very few of the important texture details of the image are lost. When the window size increases beyond 5, a blurring effect is introduced into the image, reducing the degree of contrast and making the border of the ROI unclear.

A) 3×3 window size

B) 5×5 window size

C) 7×7 window size

D) 9×9 window size

E) 11×11 window size

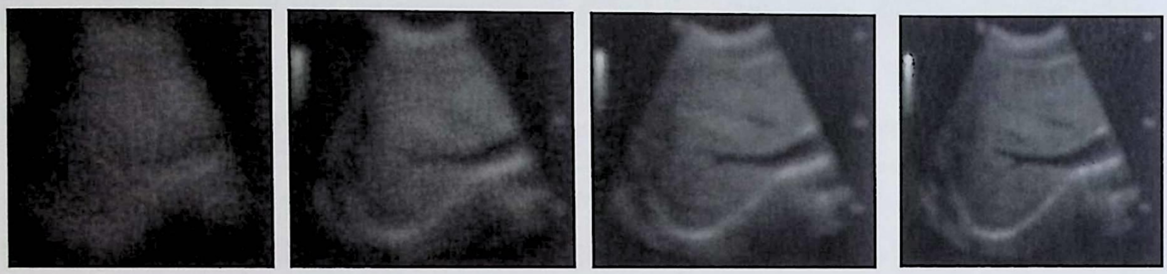
***Figure 4-5: Effects of different-sized Median Filters: A) Median 3×3 window size, B) Median 5×5 window size, C) Median 7×7 window size, D) Median 9×9 window size and E) Median 11×11 window size***

The Fourier Butterworth filter is evaluated by setting a sequence of cut-off frequencies. For this test, the cut-off frequencies were set at 10%, 20%, 30% and 40% respectively. Figure 4-2 presents the image quality metrics whereas Figure 4-6 shows the output images after applying the filter at the corresponding cut-off threshold.

In general, both the image quality metrics in Table 4-2 and the output images from each different frequency cut-off threshold shown in Figure 4-6 depict the same picture. As the cut-off threshold increases, the image quality improves, and the images are less blurry around the edges. However, the contrasts of the output images remain poor unless the cut-off frequency threshold increases further. In particular, at the 10% cut-off threshold, the ROI has disappeared into the background. At the cut-off threshold of 40%, the image has the best quality with the least noise and a more clearly identifiable ROI than the other images.

***Table 4-2: Shows the MSE, PSNR and SNR for the Fourier Butterworth filter***

| Cut off | MSE | PSNR | SNR |
|---|---|---|---|
| 10 | 0.004342 | 71.75402 | 55.40223 |
| 20 | 0.002517 | 74.12171 | 57.76992 |
| 30 | 0.001681 | 75.87434 | 59.52254 |
| 40 | 0.001235 | 77.21386 | 60.86207 |



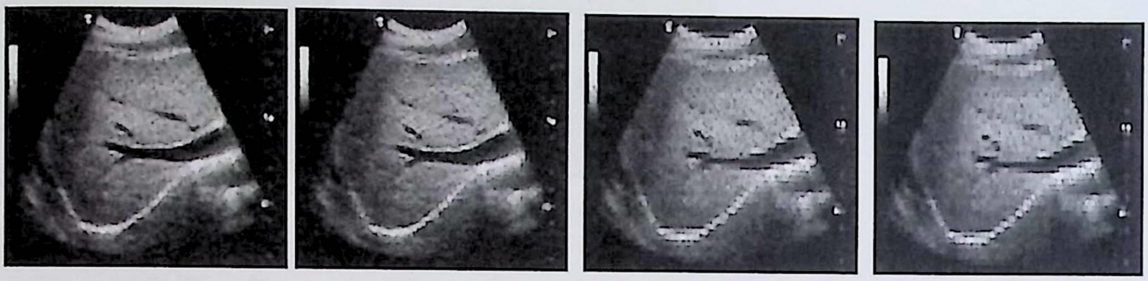
(A)Cut-off frequency 10% (B) Cut-off frequency 20% (C) Cut-off frequency 30% (D) Cut-off frequency 40%

***Figure 4-6: Effects of the Fourier Butterworth filter: A) Fourier Butterworth Filter Cut-off: 10; B) Fourier Butterworth Filter Cut-off: 20; C) Fourier Butterworth Filter Cut-off 30; D) Fourier Butterworth Filter Cut-off: 40.***

The wavelet transform approach, as a noise de-speckling technique, is based on eliminating bands in a single decomposition level with at least one high-frequency component (HL, LH and HH) removed (see Figure 4-7 and Table 4-3). Theoretically, these high-frequency bands, especially the HH band, contain noise. By closely examining the output images, we can see that in the images with HL and HH bands removed contain additional white spots that are not in the original image, which represents a deterioration of the quality metric results. On the other hand, we know that the LH band contains high frequency information representing horizontal edges, and the image tested has more vertical edges than horizontal (although most of the lines are diagonal), so if we remove this band, we remove a lot of noise but not much information relevant to any edges.

*Table 4-3: Shows the MSE, PSNR and SNR for the Wavelet Filter*

| Level | Band | MSE | PSNR | SNR |
|---|---|---|---|---|
| **1** | HL | 0.025897 | 63.9983 | 47.62165 |
| **1** | LH | 0.025068 | 64.13957 | 47.76292 |
| **1** | HH | 0.026149 | 63.95623 | 47.57959 |
| **1** | LH-HH | 0.024836 | 64.17991 | 47.80326 |



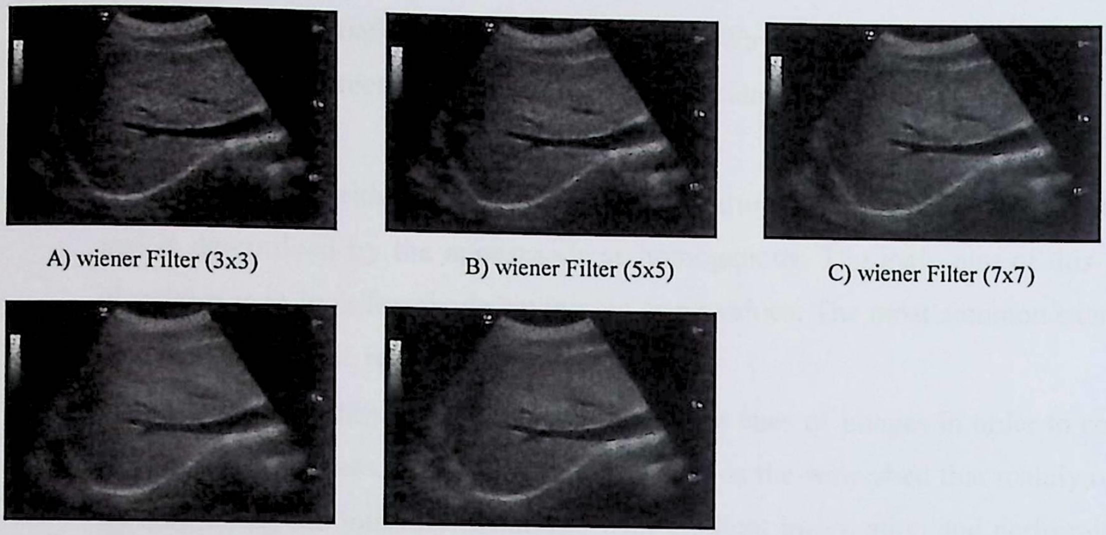
A) Wavelet Sub-band-HL B) Wavelet Sub-band-LH C) Wavelet Sub-band-HH D) Wavelet Sub-band-LH-HH

***Figure 4-7: Effect of the Wavelet Transform-Level 2: Effect of the Wavelet Transform-Level 2: A) Wavelet Sub-band-HL, B) Wavelet Sub-band-LH, C) Wavelet Sub-band-HH and D) Wavelet Sub-band-LH-HH***

The Wiener filter has also been evaluated with different window sizes (3×3, 5×5, 7×7, 9×9 and 11×11). The image quality metrics reported in Table 4-4 confirm that image quality deteriorates as window size increases in a similar pattern as that seen for the median filter described earlier; however, the amount of quality reduction appears to be much less than that for the median filter. A close visual inspection of the output images in Figure 4-8 shows that the filtered image with the (3×3) window size is of good quality, but the output images with larger window sizes still have ROIs with clear borders, although the texture patterns within the ROI are adversely affected.

*Table 4-4: Shows the MSE, PSNR and SNR for the Wiener Filter*

| window | MSE | PSNR | SNR |
|---|---|---|---|
| **3** | 0.000648 | 80.01321 | 63.63656 |
| **5** | 0.000683 | 79.78409 | 63.40745 |
| **7** | 0.000755 | 79.35194 | 62.9753 |
| **9** | 0.000828 | 78.94957 | 62.57293 |
| **11** | 0.000907 | 78.55645 | 62.17981 |



A) wiener Filter (3x3) B) wiener Filter (5x5) C) wiener Filter (7x7)

*D) wiener Filter (9x9)* *E) wiener Filter (11x11)*

***Figure 4-8: The effect of the Wiener Filter: A) Wiener Filter (3×3), B) Wiener Filter (5×5), C) Wiener Filter (7×7), D) Wiener Filter (9×9) and E) Wiener Filter (11×11).***

In a comparison of the evaluation results across all four noise filters, the Wiener filter was found to produce the best image quality, and this may help to improve diagnosis accuracy through preserving texture features. On the other hand, one can argue that the wavelet transform can help to improve the images for the purposes of segmentation purpose because it helps to highlight the edge of the ROI.

## 4.3 Intensity-based Segmentation Techniques

Traditionally, algorithms aiming to separate structures or organs of interest from other regions (background) utilise distinctive and quantifiable features. Useful features include image intensity distribution in the spatial as well as frequency domains, entropy, and gradient magnitude. The segmentation procedure aids in looking for the pixels with values within the defined ranges that are established in the pre-determined thresholds. Manual or automatic selection is an effective way to choose thresholds that used in the algorithms. In manual selection, theoretical knowledge and trial experiments that are required to determine the appropriate threshold values. Trial experiments are needed to combine information from the images and to automatically find the adaptive threshold values. The Otsu's method (N. Otsu 1979) is one of the examples commonly used in obtaining the threshold values with image histogram. Based on the information that defines threshold values, it has been confirmed that algorithms have different classification namely the edge-based, region-based, and hybrid algorithms.

- In edge-based algorithms, the threshold values relate to the edge information and are used as structured based on their appearances at the edge points. The most common examples of edge-based algorithms include the Wavelet transform, Canny edge detection (a common detection algorithms), the Laplacian edge detection, and the Sobel edge detection.

- Region-based algorithm is where quantifiable features are observed within the structure and is determined by the appearance as homogenous. The main aim of this type of algorithm is to look for pixels having the same values. The most common example of this algorithm is the region growing algorithm.
- The hybrid algorithm mainly combines different cues of images in order to complete segmentation process. One of the key examples is the watershed that mainly operates by combining the intensity of images with gradient information and performing segmentation using mathematical morphology.

An ultrasound image with empty GS and ROI is presented in Figure 4-9 (A). Figure 4-9 (B)-(F) depicts the segmentation results of region growing algorithm, watershed algorithm, Otsu threshold, Canny and Laplacian edge detections. In relation to the region growing algorithm, the GS boundary is minted well but there is over-segmentation of the ROI boundary since it leaks outside. In addition, there is over segmentation problem in the watershed algorithm that occurs as a result of artefacts and speckling noise. This over-segmentation is evident across different image areas due to various pixels with local gradient magnitude. When the images obtained from ultrasounds are affected by the speckling noise as well as intensity of homogeneity, over-segmentation also occurs and this is found on the segmentation results of threshold-based algorithms. The Canny Laplacian algorithms edge detection provides discontinuous boundaries that result from the different level of grey scale in ultrasound that affects the texture and the speckle noise. In addition, there are no representations of the spatial relationships of edge points; therefore the majority of detection boundaries are poorly connected and may appear incomplete. As indicated in Figure 4-10, similar algorithms have been applied in this work to obtain the ovarian tumour using the ultrasound image. Similar results were obtained when compared with the ultrasound images obtained through empty GS. As a consequence, the algorithms, when used alone, are ineffective in segmenting the right ROI. The researchers have examined and found that algorithms are rarely used alone. Instead, they are effective when *prior* knowledge and efficient pre-segmentation is incorporated. Thus, this work proposed an effective algorithm that can enhance threshold-based segmentation through the combination of basic segmentation solutions, sophisticated processing techniques for image, and priori domain knowledge.

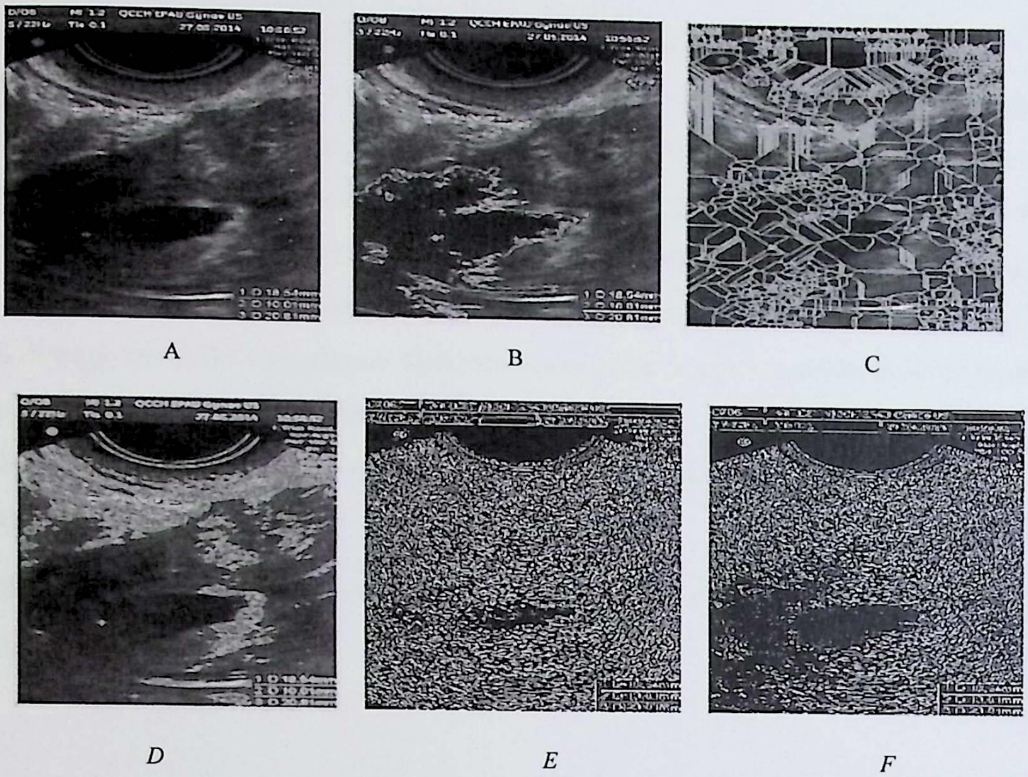



***Figure 4-9: A) Ultrasound image of a pregnancy case, B) region growing algorithm, C) watershed algorithm, D) Otsu threshold, E) Canny edge detection and F) Laplacian edge detection.***

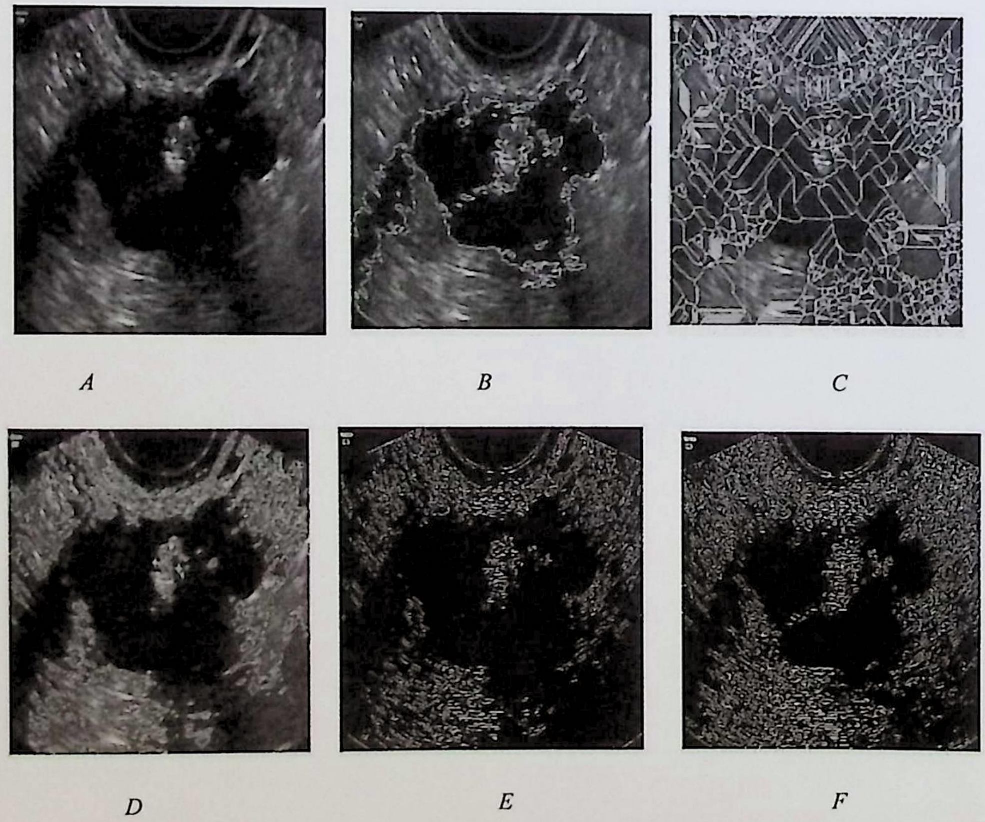



***Figure 4-10: A) Ultrasound image of ovarian tumour, B) region growing algorithm, C) watershed algorithm, D) Otsu threshold, E) Canny edge detection and F) Laplacian edge detection.***

## 4.4 Proposed Automatic Segmentation Method

Figure 4-11 outlines the processing framework behind the proposed algorithm for segmenting GS from a static B-mode image and measuring its dimensions to allow identification of cases of miscarriage. To accurately locate the GS in the image, the proposed solution first uses the wavelet transform to suppress the speckle noise by eliminating the high-frequency sub-bands and preparing an enhanced image. This is followed by an initial segmentation step that isolates the GS through the following stages: first, the mean value is used as a threshold to binarize the image. This is then followed by filtering out unwanted objects based on their geometric features according to prior domain knowledge. The region growing algorithm is then applied as a post-processing step to finally identify the correct GS.

The following subsections explain the details of each stage of the proposed automatic GS segmentation solution.

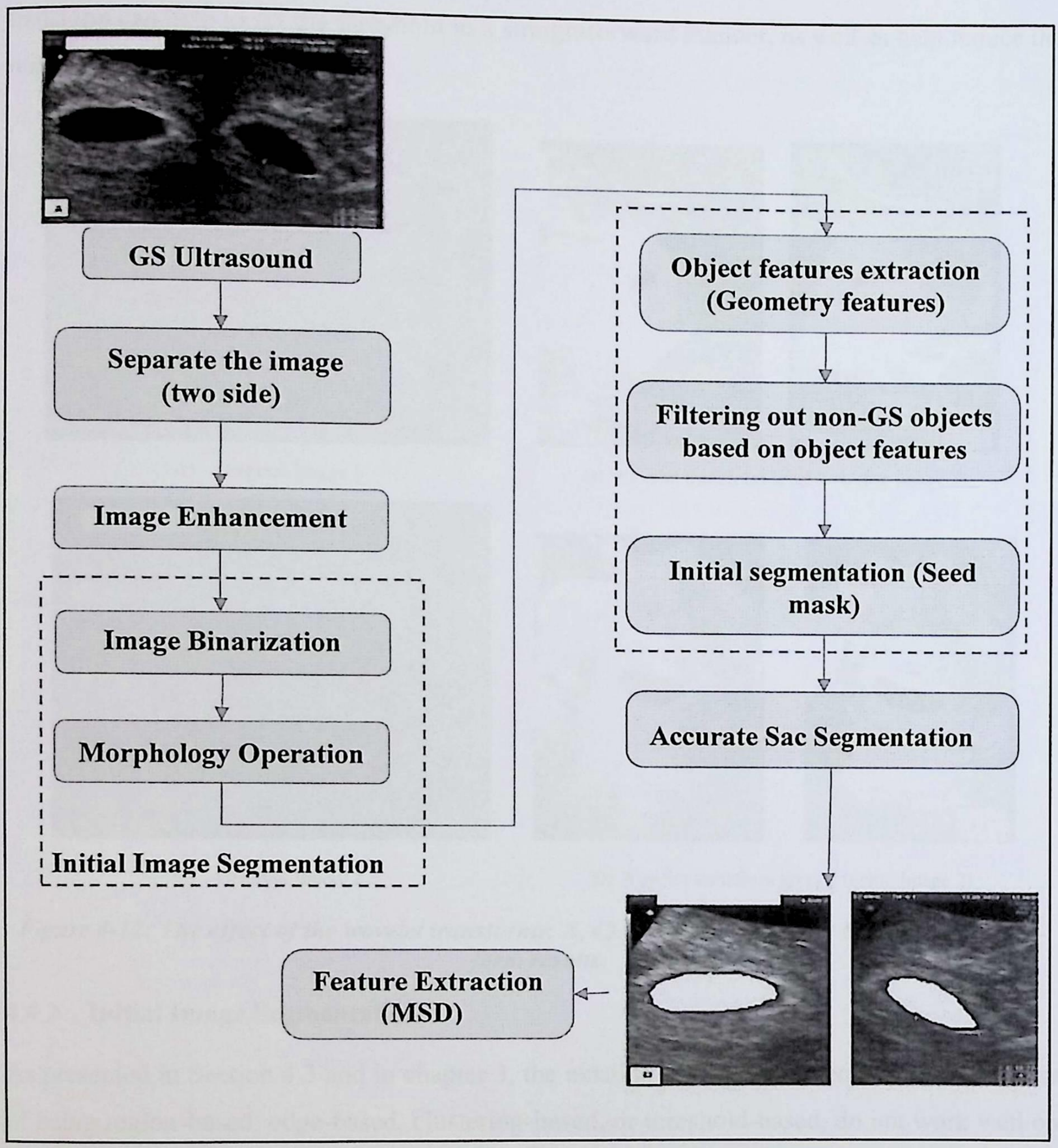
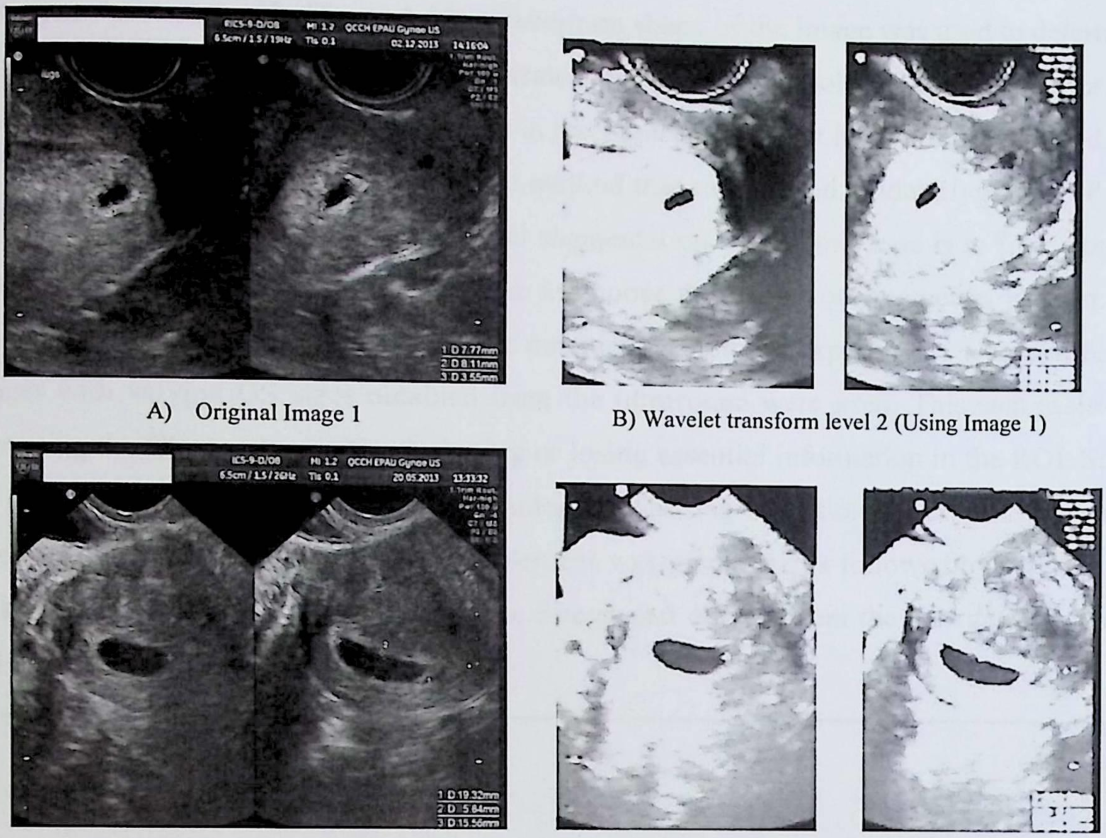



*Figure 4-11: Fully Automatic Segmentation System*

### 4.4.1 Image Enhancement

The purposes of this stage are to accurately locate structure boundaries, better visualize the positions of any structures, and to quantify geometry features. In this work, two-level Haar wavelet transform is used to suppress the speckle noise. The wavelet transform approach is based on eliminating coefficients in the HH2 sub-band (HL2, LH2 and HH2) (See section 4-6). As described previously, the HH sub-band contains noise, so, if we set the coefficients within the sub-band to zero, a lot of noise but not much edge-related information can be removed. Figure 4-12 shows two cases of GS ultrasound images. As we can see, the wavelet transform can help to set the threshold in a straightforward manner, as well as help reduce the number of FPs.



A) Original Image 1 B) Wavelet transform level 2 (Using Image 1)

C) Original Image 2 D) Wavelet transform level 2 (using Image 2)

***Figure 4-12: The effect of the wavelet transforms: A, C) original images and B, D) wavelet transform results.***

### 4.4.2 Initial Image Segmentation

As presented in Section 4.3 and in chapter 3, the existing segmentation techniques, regardless of being region-based, edge-based, clustering-based, or threshold-based, do not work well on most ultrasound images because of the huge variations in intensity values from one image to another. Therefore, we proposed a new hybrid technique based on threshold and geometric features to segment and extract the regions amongst which the GS is located.

A threshold technique was used to binarize the image because it is the simplest compared with other segmentation techniques. Nonetheless, this technique of thresholding could be the most efficient technique for segmentation, particularly in specific applications. Typically, one threshold value is used to segment an image into objects and background. However, it is common to use multiple thresholds to segment images into various regions.

Additionally, thresholds are characterized as either local or global (constant all through the image) when different thresholds are selected in accordance to the local characteristics of varying areas in the image. In Figure 4-13, a histogram shape of the image was used to determine the threshold. The figure clearly demonstrates that using threshold is not enough for the achievement of a segmented image, where in the parts of an object have multiple FP and TP. Thus, we have suggested for an appropriate method that can be used in identifying the FP and TP using the Geometry features. The initial segmentation step in this case is to filter out the non-ROI. Experimental selection was done to choose the values of parameters used in this initial stage. This selection was done after conducting multiple experiments where different images with varying GS sizes obtained from the ultrasound were used. This step assists in promoting the GS structure without altering or losing essential information in the ROI. Suing the image output from threshold, a morphological opening operation is performed where 6 pixel radius with a disk-shaped structure element was used. This is followed by carrying out background subtraction operation to isolate foreground objects from the background objects and this enhances correct detection of GS.

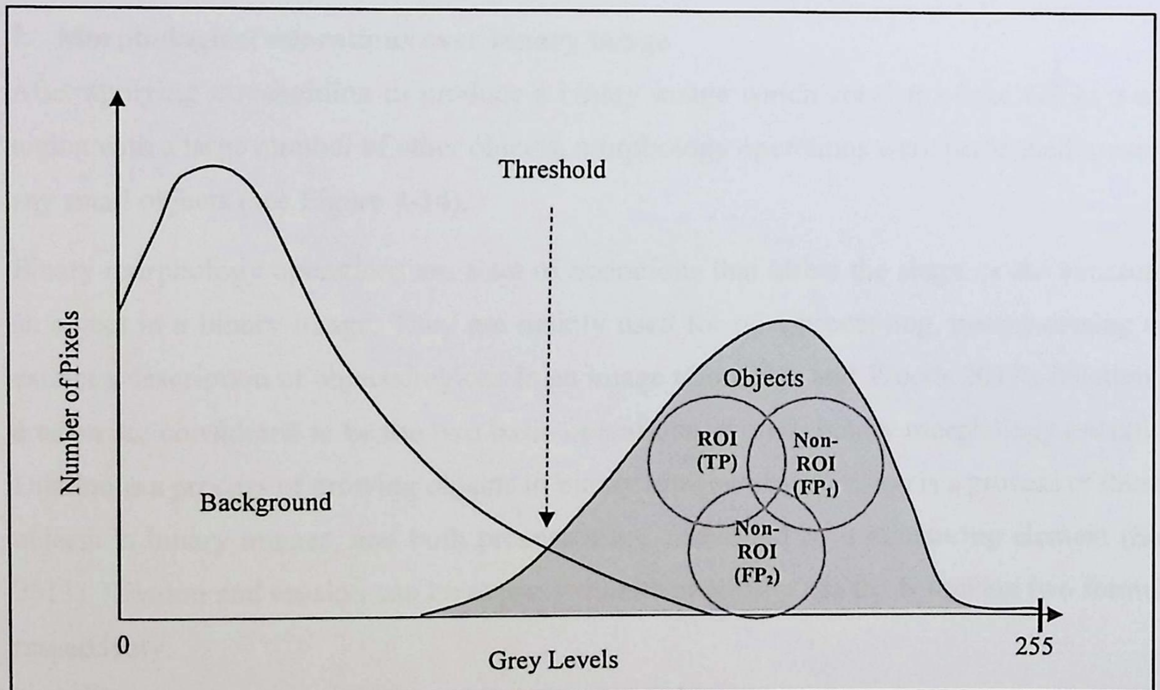
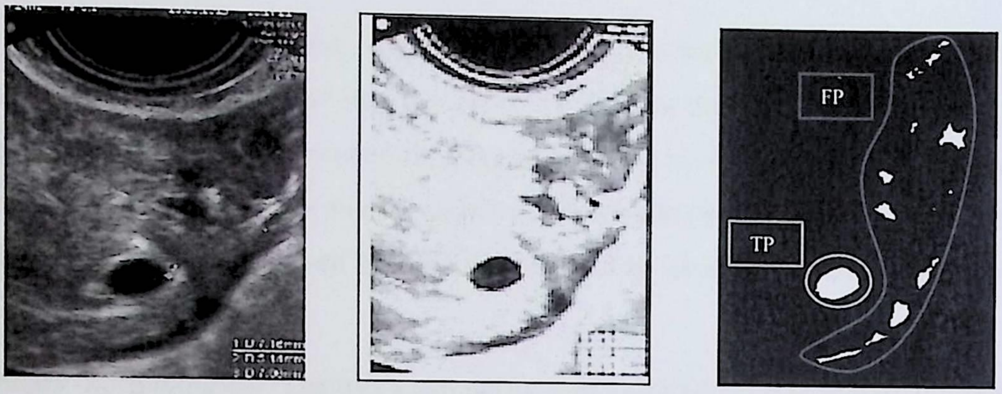



*Figure 4-13: Histogram thresholding example*

## 1. Image Binarisation

We have use a simple but effective threshold, which is the mean value of the image. This threshold is adaptive, as it changes depending on the intensity level of the image. This technique is not usually sufficient in itself to accurately extract the GS because the quality of the segmentation is critical to an accurate measure of the diameters of the GS. By analysing the threshold outputs, one can identify two limitations. Firstly, each image has a number of binary objects near the GS (see Figure 4-14). Secondly, the binary image occasionally does not contain the entire area of the GS. Therefore, furthersteps described in the following subsections are proposed to address these limitations.




A) Original image B) Enhanced image C) Segmented image

*Figure 4-14: The FP and TP based on threshold: A) original image - GS ultrasound image B) enhanced image using three-level wavelet and C) segmented image (includes FP and TP)*

**2. Morphological operations over binary image**

After applying thresholding to produce a binary image which consists of the GS as a white region with a large number of other objects, morphology operations were performed to remove any small objects (see Figure 4-14).

Binary morphology operations are a set of operations that affect the shape or the structure of an object in a binary image. They are mainly used for post-processing, pre-processing or to extract a description of objects/regions in an image (Gonzalez and Woods 2017). Dilation and erosion are considered to be the two basic operations of other binary morphology operations. Dilation is a process of growing objects in binary images while erosion is a process of thinning objects in binary images, and both processes are controlled by a structuring element (Soille 2013). Dilation and erosion can be expressed mathematically via the following two formulae, respectively:

$$A \oplus B = \left\{ z \mid (\hat{B})_z \cap A \neq \emptyset \right\} \qquad (4\text{-}4)$$

$$A \ominus B = \{ z \mid (B)_z \cap A^c \neq \emptyset \} \qquad (4\text{-}5)$$

Where A is the binary image, $A^c$ is the complement of A and $(\widehat{B})_z$ is the structure element B after reflection and translation by z. The dilation of A by B is defined as the set of all displacements z where A and $\widehat{B}$ overlap by at least one element. The erosion of image A by structure element B is the set of all structuring element origin locations where the translated B has no overlap with the background of image A.

Two important morphology operations, opening and closing, are defined by combining dilation and erosion. Opening is an operation that has ability to smooth the contour of an object and reduces the presence of some small unwanted objects. Opening is simply implemented by erosion followed by dilation. Closing is an operation that closes narrow gaps and fills small holes to smooth objects. Closing is achieved by dilation followed by erosion (Gonzalez and Woods 2017).

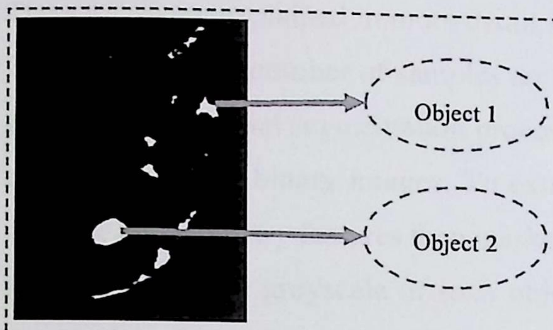
In general, binary morphological operations can perform in a similar manner to spatial filtering. They can be useful in a pre- and post-processing of a segmented binary mask. Two precautions should be considered when using such operations. The first one is that they need careful setting of their parameters (i.e., size and orientation of structure elements in open and close operations) to give the required result. Furthermore, besides their intended enhancements to certain objects in the image, other objects might inadvertently be negatively affected by applying those operations blindly to the whole image.

The values of the parameters used in this stage were selected empirically after a number of experiments had been carried out using different ultrasound images with different GS sizes, taking into account the enhancement of the structure of the GS without any simultaneous loss of important information in the specific ROI. The morphological opening operation is carried out on the image output from the threshold, using a disk-shaped structure element of a six-pixel radius, followed by a background subtraction operation to separate out foreground objects from the background and detect the GS correctly.

**3- Filtering our non-Sac objects based on geometry features**

The central problem that needs to be addressed at this stage is the fact that upon completion of the segmentation, irrelevant binary objects (i.e., objects that don't not represent the GS) still exist in the binarized image. Therefore, it is necessary for a fully automated system to identify these objects and consequently remove them. This process of filtration is performed based on the geometric characteristics of such non-sac objects.

Our object detection procedure classifies image objects based on their geometric features. The aim here is to use powerful geometry features to reduce the overlap between the false positives and true positive. These features support the classifier model in the identification of the correct ROI. In this model, we have used the following features: object circularity, object area, object solidity, and object mean greyscale value (see Figure 4-15). These features are described as follows:




| Features | Object 1 | Object 2 |
|---|---|---|
| Circularity | × | √ |
| Object area | × | √ |
| Solidity | × | √ |
| Object mean of grayscale value | × | √ |

*Figure 4-15: The features of the filtering out non-sac object step*

- **Circularity.** The circularity of an object is the ratio of the square of the perimeter to the area:

$$\text{Circularity}=\frac{\text{Per}^2}{4*\text{pi}*\text{area}} \qquad (4\text{-}6)$$

Where the perimeter (Per) is the distance around the inner boundary of an object and the area is the number of pixels inside the object. Because a sac is roughly spherical in shape with roughly circular projections in ultrasound images and also have a little circularity value region, sac circularity is a good indicator of the likelihood of whether a binary object is a true sac or not.

- **Object area.** Object area is defined as the number of pixels making up an object. The object area allows us to distinguish between different regions including sacs, even if they are small.
- **Solidity:** Solidity measures return a scalar specifying the proportion of the area of a region in pixels to the convex hull area that contains the region. Solidity is computed as:

$$\text{Solidity}=\frac{\text{Area}}{\text{Convex Area}} \qquad (4\text{-}7)$$

- **Object mean of grayscale value.** The mean value of the object is the average of each object based on its grayscale value. The mean of the sac must have a low greyscale value near to zero (black) whereas other non-sac objects in ultrasound images may have higher greyscale values due to the presence of greyscale texture patterns within these objects. This is a good diagnostic of having correctly distinguished the true sac from any non-sac objects.

The proposed object detection model involves two main stages: first, the training stage, to build a model based on the training samples; and second, the testing stage to classify the objects. These stages are explained in more detail in the following subsections.

**Training Stage:** a number of samples are used to train the SVM classifier. We have used the previous stage (initial segmentation process) to binarize the image. Several objects have been extracted from the binary images. To extract the features, we have used the binary mask to calculate the geometry features then masked the original image to calculate the solidity and the mean value for the greyscale of each objects. Somewhat subjectively, we have labelled the features into the positive class and negative class by looking into the objects. Finally, we fed the extracted features and labels into the SVM classifier (See Figure 4-16).

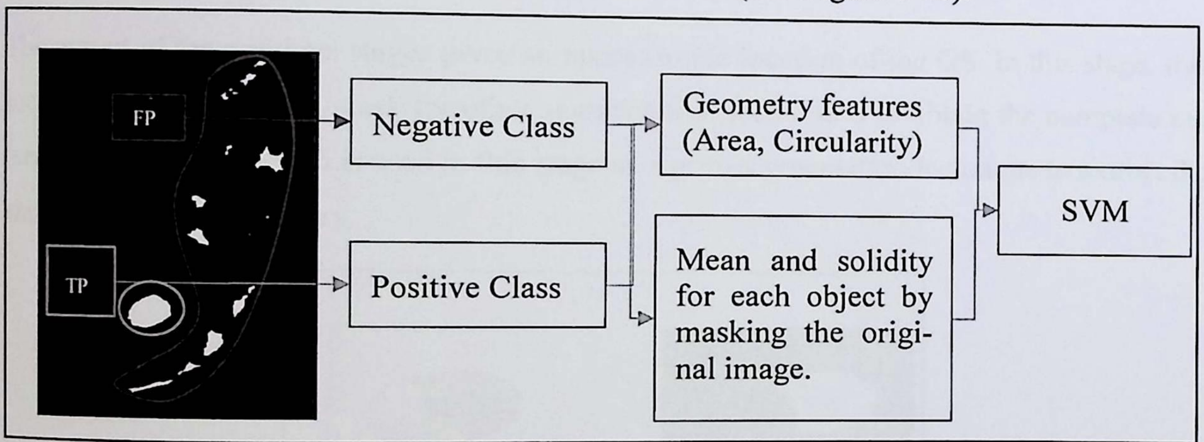



*Figure 4-16: Training stage for filtering out non-GS*

**Testing stage:** the output of the initial segmentation stage is binary image include number of objects (TP and FP). The aim of this stage is to identify the TP object. To achieve this, our proposed model scans the image object by object. For each object, four features (circularity, area, mean and solidity) are extracted. By using the score of the trained SVM model the correct GS is then identified.

Figure 4-17 illustrates the resulting images from each intermediate step in this initial GS segmentation stage.

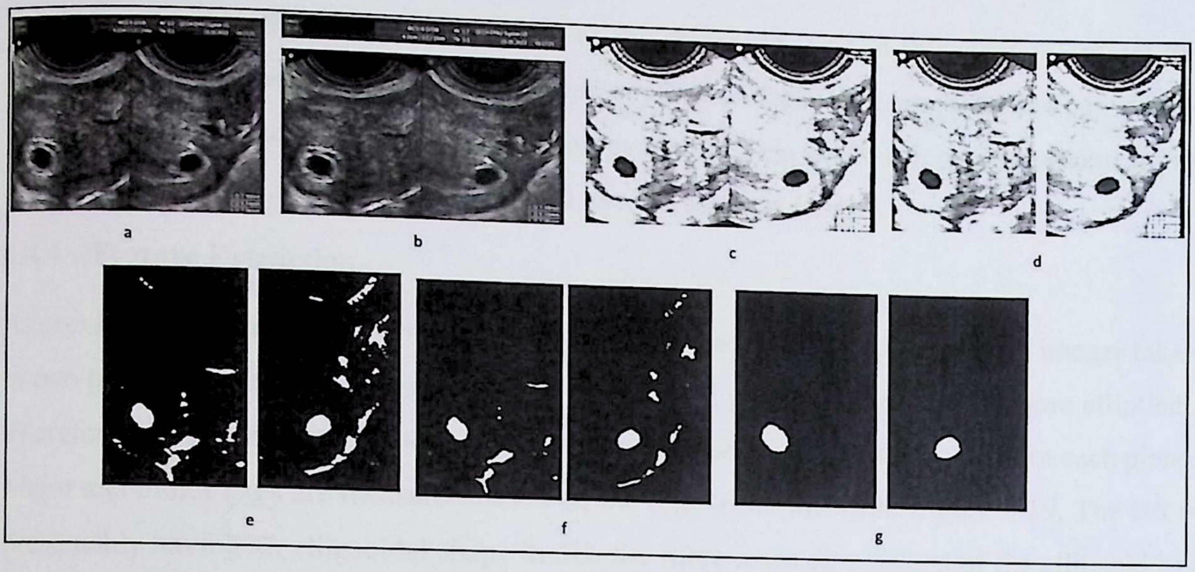



*Figure 4-17: GS segmentation steps: A) original image, B) cropped 70 pixels, C) enhanced image, D) separated both sides, E) threshold output, F) clean image (morphology operations) and G)filtering out non-GS objects*

### 4.4.3 Accurate Sac Segmentation

The output of the previous stages gives an approximate location of the GS. In this stage, this location is used as a seed mask for other segmentation techniques to obtain the complete sac (see Figure 4-18). The RG is used in this stage as a post-segmentation technique to extract the sac with a correct boundary.

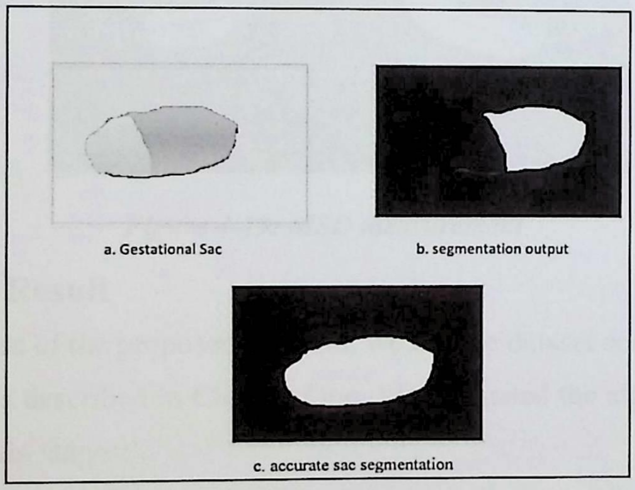



*Figure 4-18: Accurate sac segmentation output*

**Region Growing (RG):** the output of the filtering out the non-GS stage is an initial mask. This mask was used as a start point for the RG. In this stage, a threshold classifies a given pixel based on the grayscale value of the starting mask with the neighbouring pixels to check the difference. This difference is used as a value to determine whether the neighbouring pixels should be added to the region. The values of the threshold used in this stage to extract the complete ROI were selected empirically after a number of experiments had been carried out using different ultrasound images with different texture GSs without losing important information from the ROI. The optimal threshold used in this stage is 0.4.

### 4.4.4 Feature Extraction

As previously explained, each image we processed in this study contained two sac images taken in two planes (sagittal and transverse). The GS in each plane usually appears more elliptical. Therefore, our system aims to locate the best-fitted ellipse for the segmented GS for each plane. Major and minor axes are measured based on the ellipse, as shown in Figure 4-19. The GS is presumably having an ellipsoidal shape in 3D, the three main dimensions of the ellipsoid has generally been calculated by the minor and major dimension from the sagittal plane and the major dimension from the transverse plane. The average of the three Dimension is taken as the MSD (Abdallah, Daemen and Kirk, et al. 2011).

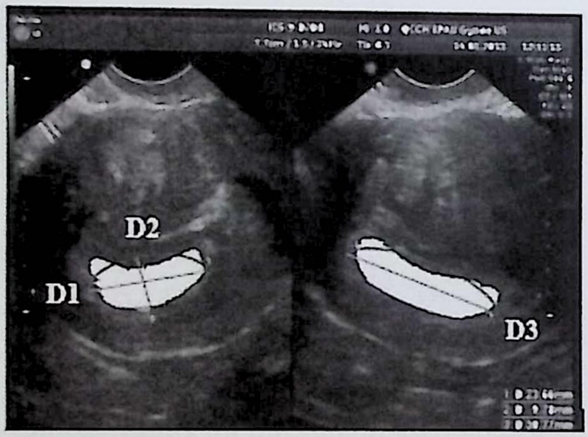



*Figure 4-19: MSD measurement*

## 4.5 Experiments Result

To test the effectiveness of the proposed solution, we use the dataset containing B-mode ultrasound images of GS as described in Chapter One. We evaluated the algorithm's performance at each of the three main stages:

- Assess the accuracy of the segmentation
- Compare the automatic measurements with the manual ones generated by a domain expert.
- Evaluate the accuracy in identifying the miscarriage cases using the MSD feature calculated from the automatic GS size measurements.

### 4.5.1 Segmentation Result

The Human Visual System (HVS) is used to manually calculate the percentage of successful segmentation of the GS. This is because the size of the available dataset is not particularly large. In the binarization stage of the solution, we found that the proposed solution successfully located the objects, including GS, in 151 out of 184 images. Therefore, the rate at which the true GS is located is around 82%.

In the stage in which non-ROI (FP) were filtered out, we found that 137 out of 151 images GSs accurately detected the TP and 14 images were unsuccessful. Therefore, the GS segmentation rate for this method is around 90 % (see Table 4-5). To evaluate the proposed model out of the whole dataset, 137 images out of 184 have segmented correctly, and only these images will use to evaluate the accuracy in identifying the miscarriage cases.

***Table 4-5: Accuracy of the threshold-based segmentation***

| | **Proposed method** |
|---|---|
| **Located the TP besides the FP** | 151/184 |
| **Accuracy** | 82.06% |
| **Success out of 151** | 137/151 |
| **Accuracy** | 90.7% |
| **Successful segmentation out of the whole dataset** | 137/184 |
| **Accuracy** | 74.4%. |

Figure 4-20 shows some example results for threshold-based segmentation where the first column presents the original image, the second column presents the initial segmented GS and other FP objects, and third column presents the true GS after filtering out non-ROI objects. For illustrative purposes, the results of segmentation for the second and third columns are superimposed in green lines on the original image. Figure 4-20-A illustrates that the proposed method segmented the true GS together with a number of FP objects. The filtering out of non-ROI also succeeded because of the regularity of the shape of the GS in terms of its circularity compared to FP objects. Figure 4-20-B illustrates a difficult case, i.e., an ultrasound image with a very small GS with an irregular shape. In this case, since the initial segmentation results in too many FP objects of various shapes, the threshold-based checks applied by the post-processing stage of the proposed algorithm has failed to recognise the true ROI. Figure 4-20-C shows a case where the GS has a poor and undistinguished border, where the proposed segmentation algorithm cannot catch the proper border of the ROI, such that the segmented region contains irrelevant parts of the background. Figure 4-20-D shows another example where the proposed method does not even catch the ROI in the initial segmentation stage, and consequently fails to recognise the true GS. These examples not only illustrate the strengths and limitations of the proposed solution, but also highlight the difficulty faced by segmenting ROIs in ultrasound images.

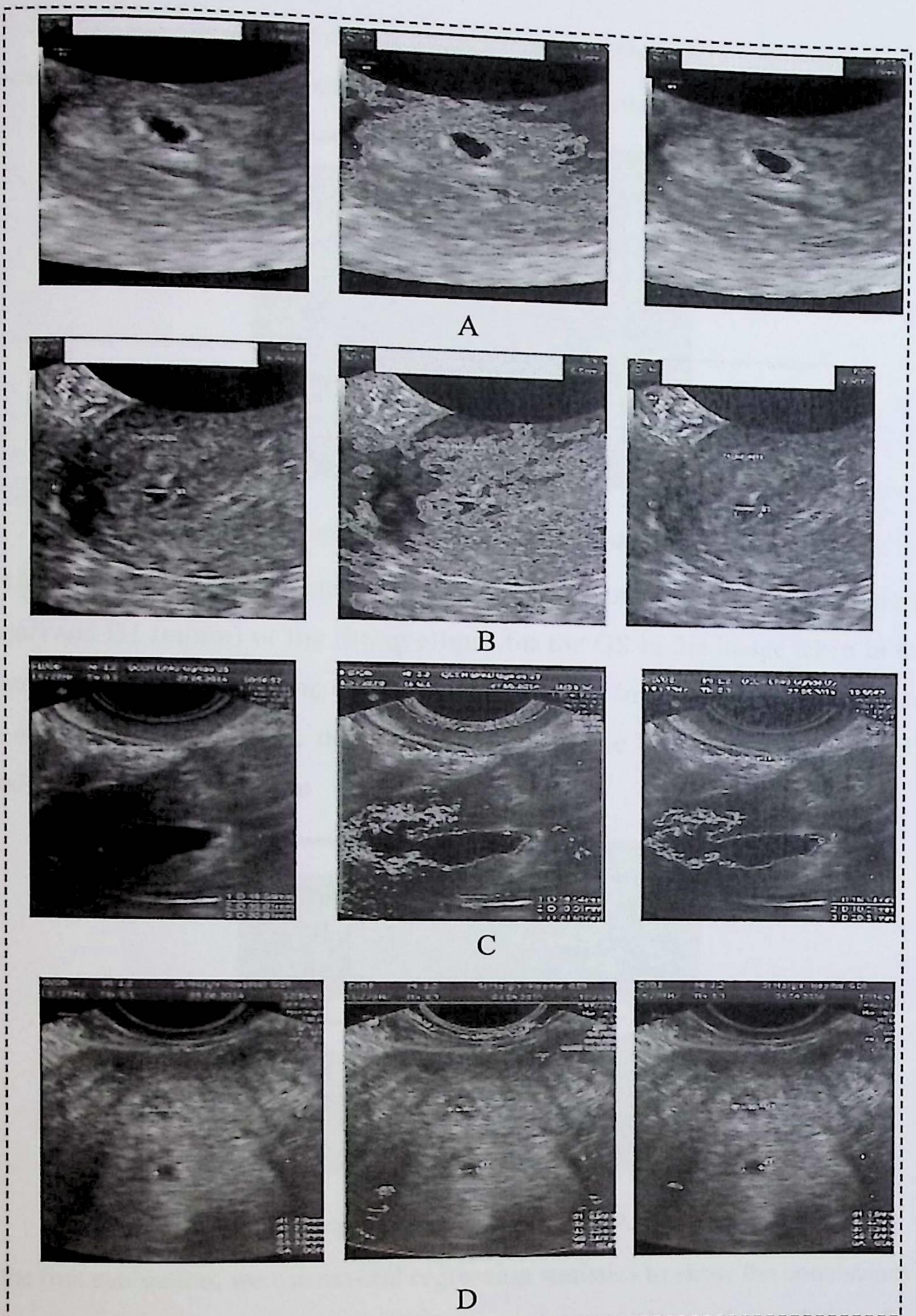


Figure 4-20: The results of trainable and threshold-based methods

### 4.5.2 Comparing Closeness between Manual and Automatic MSD Measurements

To evaluate the effectiveness of the proposed segmentation algorithm, the automatic measurements of the GS size, i.e., diameters D1, D2 and D3 are compared with the equivalent manual measurements as taken by a domain expert and provided in the ground truth document (See

Figure 4-21, read box present the expert measurements). The MSD calculated from the automatic measurements is also compared with the MSD calculated from the manual measurements.

*Figure 4-21: The manual measurement inside the red rectangle.*

Figure 4-22 shows the three diameters obtained from the best-fit ellipse to the GS in two planes: D1 (major) and D2 (minor) of the fitting ellipse for the GS in the image taken in the sagittal plane, and D3 (major) of the fitting ellipse for the GS in the image taken in the transverse plane. For the experiment, we will use the 137 images that have been segmented correctly by the proposed segmentation algorithm

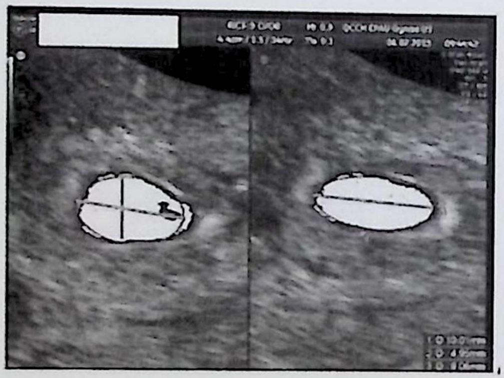


*Figure 4-22: Pregnancy images with MSD measurements.*

As for the first evaluation, we use several regression statistics to show the consistency between the manual and the automatic measurements through regression and correlation. Figure 4-23 shows the scatter plots for the automated and the manual measurements for D1, D2, D3 and MSD, respectively. The angle of the regression line is close to 45°, the value of $R^2$ and the values of other indicators in Table 4-6 are very high. All the visualisations and indicators show the level of consistency between the two kinds of measurements is high. By closely examining the figures in Table 4-6 and the scatter plots, we notice that D2 and D3 show larger variation between the manual and the automatic measurements than D1, for two reasons. First, the irregularity of some GS shapes means that the ellipse will not fit the GS precisely, particularly for

D2. This can be seen in Figure 4-24 (A) and (B) where there is a clear difference between the best-fitting ellipse and the actual shape of the GS, and hence a greater difference between the manual and the automatic measurements. Second, the manual measurements of D3 on the transverse plane for the majority of images from the ground truth document refers to the major diameter of the GS shape in that plane, and hence our automatic measurement of D3 has followed the majority convention (as shown in Figure 4-24 (D)). However, for some images in the datasets, the minor diameters have been taken as D3 (See Figure 4-24 (C)), though without any clear reason. For these images, the automatic measurements of D3 have larger differences with the manual measurements. It is interesting to note that when we compare the MSDs calculated using the manual and the automatic measurements, respectively, the difference between them is generally much less than those found for the individual diameters. This means that the MSD obtained from automatic measurements of the GS size could still be diagnostic of the empty GS suggesting an instance of miscarriage (see next subsection).

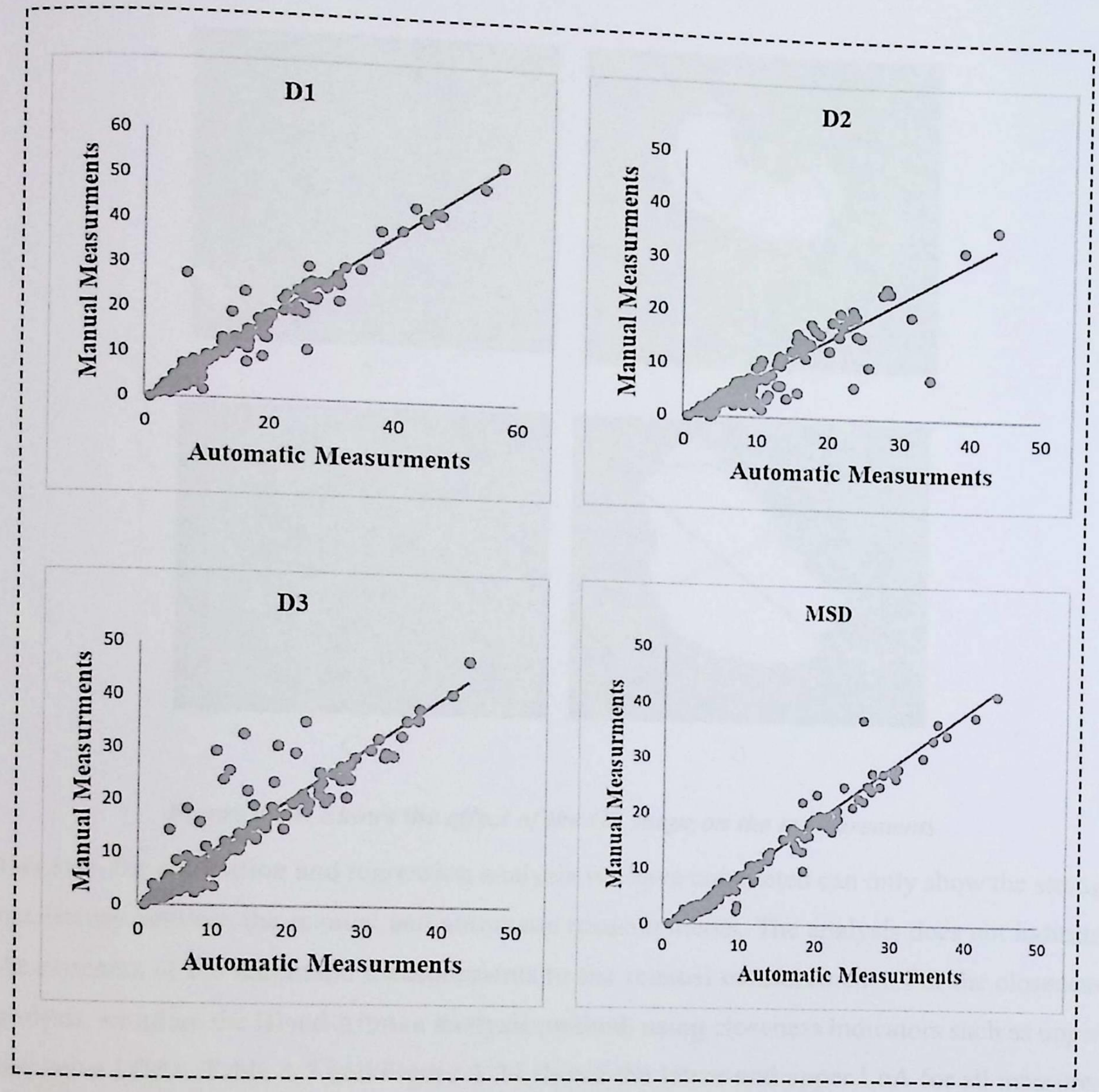



*Figure 4-23: Comparison of manual and automatic MSD measurements for both D1, D2, D3 and MS.*

*Table 4-6: Shows the regression statistics for D1, D2, D3 and MSD*

| *Regression Statistics* | | | | |
|---|---|---|---|---|
| | D1 | D2 | D3 | MSD |
| **Multiple R** | 0.963189 | 0.906607 | 0.911783 | 0.979589 |
| **R Square** | 0.927733 | 0.821936 | 0.831347 | 0.959594 |
| **Adjusted R Square** | 0.927198 | 0.820617 | 0.830098 | 0.959295 |
| **Standard Error** | 3.185926 | 3.417688 | 4.021005 | 1.858828 |
| **Observations** | 137 | 137 | 137 | 137 |

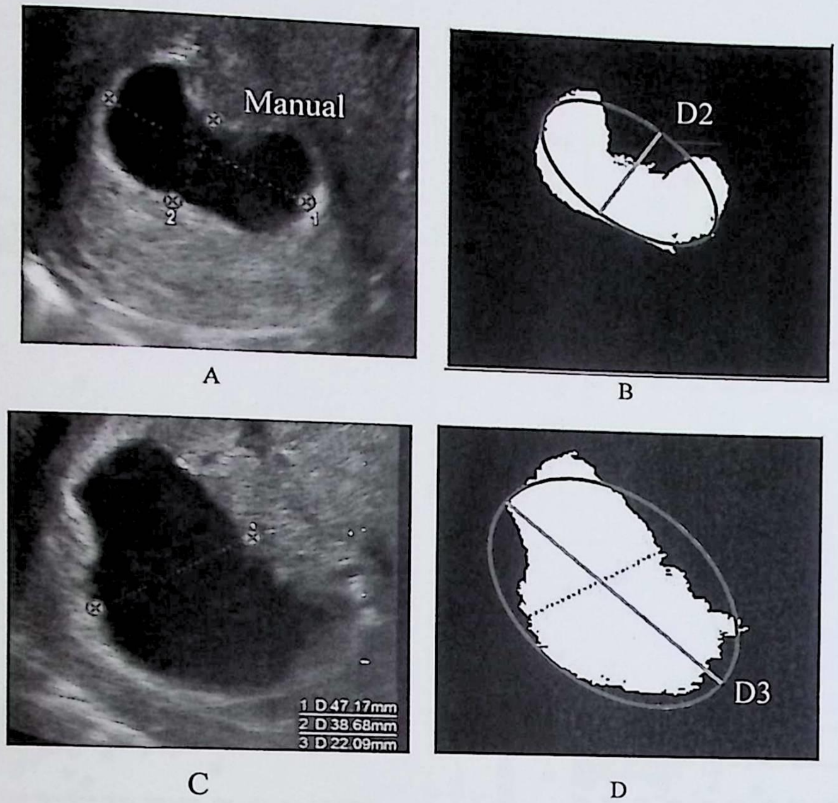



***Figure 4-24: Shows the effect of the GS shape on the measurements***

However, the correlation and regression analysis we have conducted can only show the strong consistency between the manual and automatic measurements. The analysis does not indicate the closeness of the automatic measurements to the manual measurements. For the closeness analysis, we adopt the Bland Altman analysis method, using closeness indicators such as upper and lower LOAs. Table 4-7 and Figure 4-24 shows the lower and upper LoA for all measurements.

***Table 4-7: Upper and Lower LOA for the Bland Altman analysis for D1, D2, D3 and MSD***

| | D1 | D2 | D3 | MSD |
|---|---|---|---|---|
| **Upper LOA** | 7.591 | 9.441 | 8.160 | 4.951 |
| **Lower LOA** | -4.855 | -4.106 | -8.613 | -2.414 |

The scatter plots in Figure 4-25 clearly show that the majority of measurement differences are well within the lower and upper LoA; only a few measurement differences are outside the boundary areas. Among the measurements, D1 and MSD measures have the smallest bounding area between the lower and upper LoA, indicating strong closeness between the manual and automatic measurements for these two size parameters.

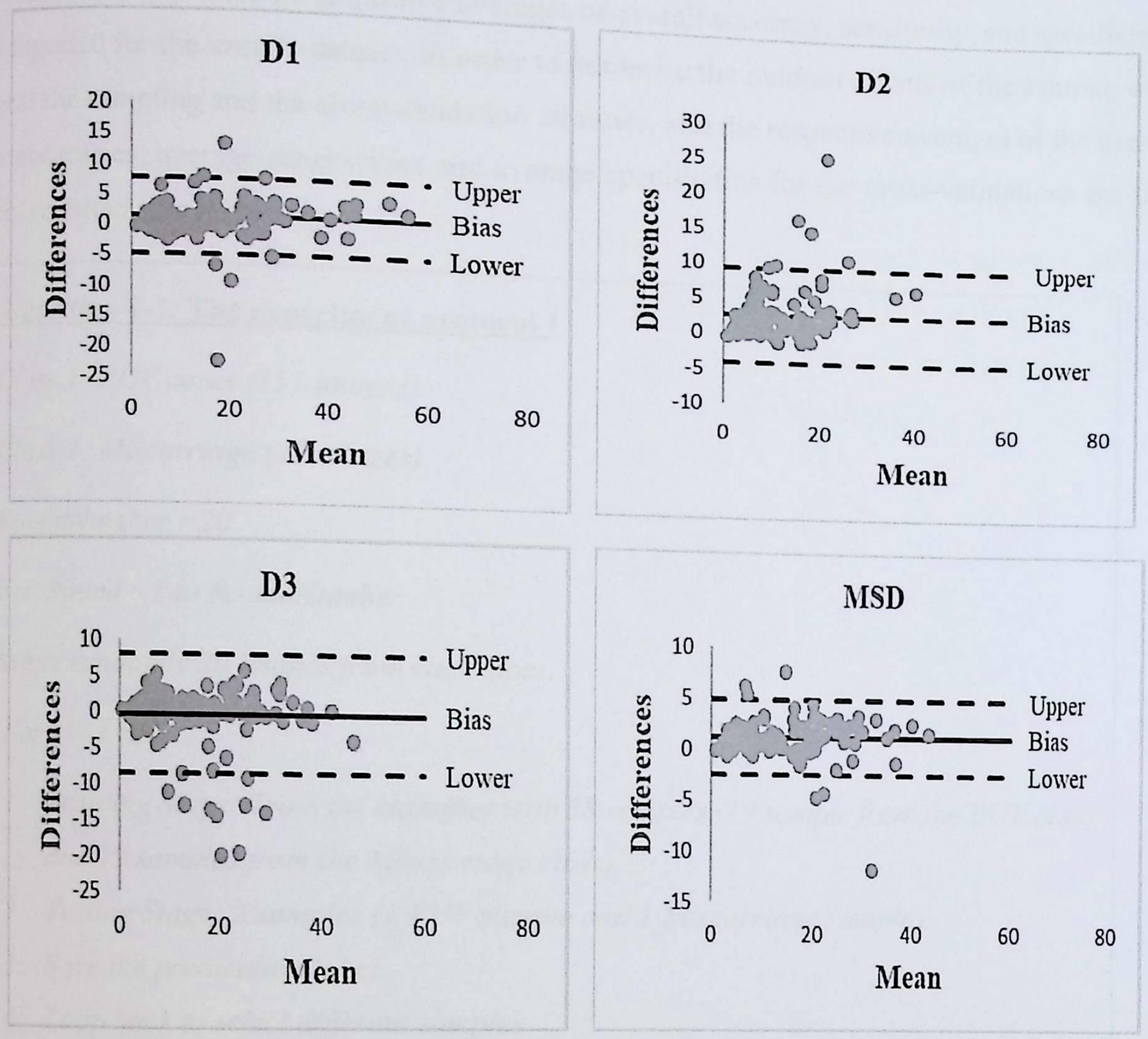



*Figure 4-25: Altman analysis for manual versus automatic segmentation and measurements for D1, D2, D3 and MS.*

### 4.5.3 Classification Results

Besides comparing the closeness of the two kinds of measurements, we also evaluate the effectiveness of the proposed segmentation algorithm through classification; namely, we want to understand if the automatic measurements of the GS based on automatic segmentation can also lead to good classification results if a GS indicates a miscarriage or PUV. We conduct two experiments: in the first, we gave PUV and miscarriage cases an equal weight in the training set using a sampling policy that is explained as follows. In each round, we randomly selected 20 PUV images and 20 images of miscarriage cases to form a random sample of examples. For each random sample dataset of 40 images, we conduct stratified leave-one-out cross-validation, i.e., at each iteration of the cross-validation, we take two images (one from each class) for testing, and the remaining 38 images are used for training. Once the leave-one-out cross validation is completed, the respective averages of overall accuracy, sensitivity, and specificity are reported for the sample dataset. In order to minimise the random effects of the sample, we repeat the sampling and the cross-validation 20 times, and the respective averages of the average accuracies, average sensitivities and average specificities for the cross-validations are finally reported (see Algorithm 4-1).

***Algorithm 4-1:* The experiment protocol 1**

*Class 1: PUV cases (157 images)*

*Class 2: Miscarriage (27 images)*

*RoundNumber =20*

*For Round= 1 to RoundNumber*

*Select randomly 20 images from each class.*

*For I = 1 :20*

1 *Training Stage: Train the classifier with 38 samples (19 sample from the PUV class and 19 samples from the Miscarriage class).*
2 *Testing Stage: 2 samples (1 PUV sample and 1 Miscarriage sample).*
3 *Save the predication label.*
4 *Loop back to select different samples*

*End I*

*Calculate the Accuracy, Sensitivity and specificity for each round*

*End Round*

*Calculate the Accuracy, Sensitivity and specificity for the whole experiment*

In the second experiment, we used a single split testing strategy instead of cross-validation by randomly selecting 33% from each class (PUV and miscarriage) in the dataset to form a testing sample, whilst the remaining 66% of each class form a training sample. This was then repeated for 50 rounds and the average accuracy, sensitivity, and specificity were reported. The reason for using this testing strategy is to follow a general practice of testing models in medical and clinical research.

The SVM classifier is used because the data objects have just one dimension, i.e., the MSD features. The SVM classifier determines the class of a testing image by calculating the absolute difference between the testing image's MSD and that of each example in the training set.

Figure 4-26 shows that an overall average accuracy of 98.70% with a sensitivity (miscarriage) of 100% and a specificity (PUV) of 94% in the first experiment. By contrast, an overall accuracy of 99.89% with a sensitivity of 100% and a specificity of 99.87% was achieved in the second experiment, as shown in Figure 4-26. The results confirm that giving a higher weight to the PUV cases compared to the miscarriage cases could give a diagnostic model that was closer to clinical practice.


| | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| First Experment | 98.7 | 100 | 94 |
| Second Experment | 99.89 | 100 | 99.87 |

***Figure 4-26: Comparison of miscarriage classification accuracy, sensitivity and specificity based on MSD.***

## 4.6 Discussion

In this section, we present different discussion areas related to the proposed solutions. **First**, in Section 4.4.1, the wavelet transform used as a filter to highlight the ROI and make the threshold method suitable for the ultrasound image segmentation. This filter was chosen after the investigation and evaluations for the speckle noise filters in Section 4.2. Based on this investigation, we argued that the wavelet transform would remove a lot of information that might be important for other tasks, i.e., texture feature extraction. However, for the segmentation task we found the wavelet filter useful because it makes the edge of the GS clear and reduced the overlap between the texture of the ROI and the background. Different wavelet levels have been tested to establish in which level of decomposition we can remove the speckle noise, highlight the ROI and still maintain the GS image. As shown in Figure 4-27, when the GS is small. We can see that the best resulting image is generated at level 2 where the contrast between the sac and background is very high. At level 4, however, we noticed that the GS image had almost disappeared. In addition, we tested the filter by eliminating the high-frequency bands (HH, HL, LH and HH&HL). Figure 4-27 shows that the elimination of any band can have the same effect on the images. Similar observations have been made for other images.

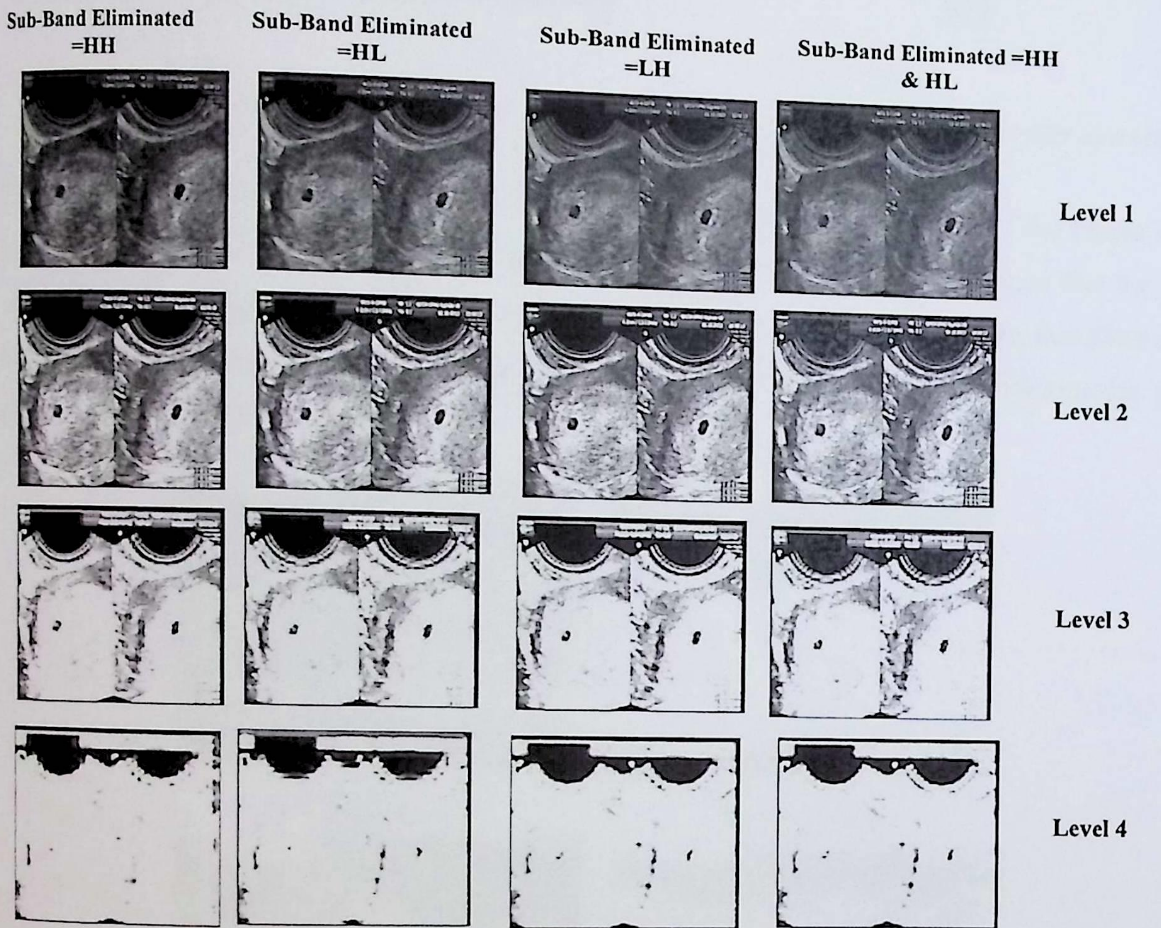


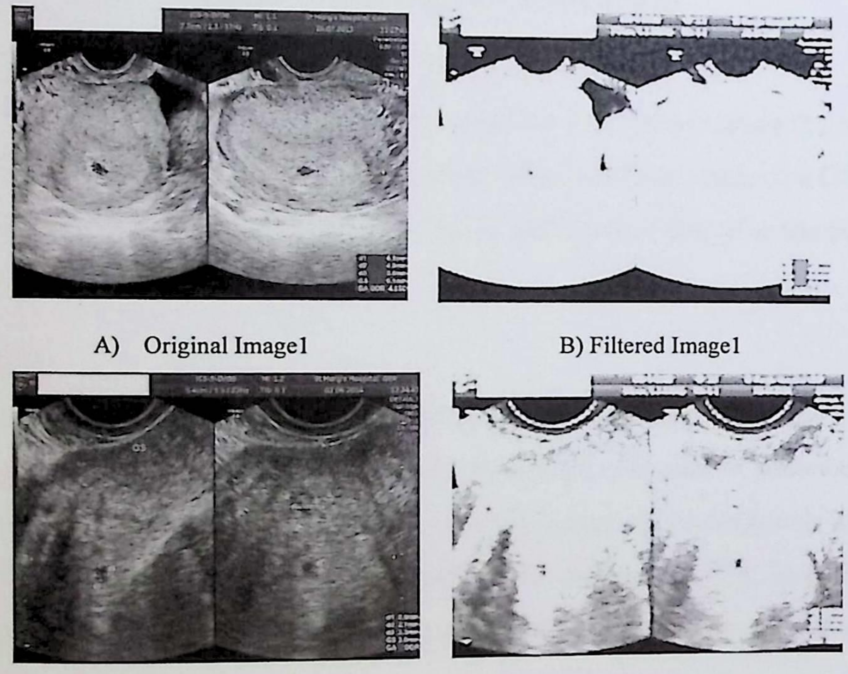

*Figure 4-27: Evaluating the wavelet transform with different levels*

Therefore, the level of the wavelet transform has an effect on the result of the segmentation more than the bands. In our dataset, we have two cases: 1) small GS special cases that were scanned in the first two weeks; and 2) a large GS. In case of a small object, increasing the wavelet level to more than 3 will make the GS disappear. With a large GS increasing the level will help because it makes the GS clear by reducing the FPs, mainly because it will make the background texture smooth. Figure 4-28 shows the application of the wavelet filter at level 3 on an ultrasound image with a large GS. We can clearly see how the wavelet transform can highlight the ROI.

A) Original image B) Filtered image

*Figure 4-28: The effect of the four-level wavelet transforms when eliminating the HH and HL bands*

Figure 4-29 shows the wavelet transform for two GS images. The first part of the figure is a blurry image with a small object. By applying the wavelet transforms, we found that the TP object disappears, though in part 2 the small object disappears as well. Therefore, that filter can help with images scanned after week 2 of the pregnancy because the GS has sufficient size and clarity to be distinguished.



A) Original Image1 B) Filtered Image1

C) Original image2 D) Filtered Image3

*Figure 4-29: The effect of the wavelet transforms: A,C) original images, and B, D) filtered images*

**Second**, the threshold was applied to the filtered image to obtain the binary image. Our objective at this stage is to separate the TP object (GS) from the binary objects. Therefore, the previous stage defiantly will affect the threshold result because if the GS disappear in the previous stage that's mean the threshold defiantly will not capture the GS. In addition, the overlap problem between the objects is one of the limitations we faced in this stage, where in some cases, the threshold cannot capture the poor GS edge.

**Third**, the filtering out non-GS objects stage has some limitations related to the geometry features, i.e., the circularity feature might be suitable for most GS images, but for some cases the circularity feature cannot capture the GS due to the irregularity of its shape, as shown in Figure 4-30.

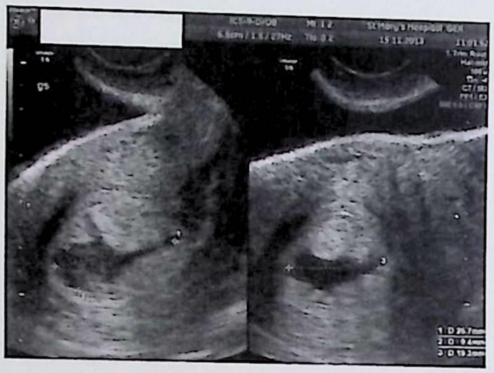


*Figure 4-30: An irregular GS*

**Finally**, region growing was used as a post-segmentation method to reduce the overlap problem. The limitations of this algorithm are the threshold used to label the pixels as a GS (or otherwise) as well as the criteria (pixel intensity) used being insufficient to describe the pixels due to the variance in the GS intensity.

## 4.7 Summary

The aim of this study was to explain the challenges faced in order to enhance and segment the ultrasound images, followed by a review of the existing methods used to enhance the ultrasound and identify the best methods by which to improve the accuracy of diagnosis as well as determine the best filter to highlight the ROI and reduce the number of FPs. In addition, the more traditional segmentation methods were evaluated by application to the ultrasound images.

This chapter investigated the closely related issues of speckle noise, segmentation and the classification accuracy of automatic identification of miscarriage cases using ultrasound imaging systems. The aim of this chapter was to understand the effect of speckle noise on automatic segmentation, diagnoses, and to use the knowledge so acquired to design a fully automatic framework for accurate segmentation and classification of early miscarriage cases. The segmentation was divided into two stages. Firstly, the initial segmentation of the sac was performed after removing unwanted objects based on geometric features. Secondly, a refinement stage was used to accurately identify the entire sac by applying the region growing algorithm that allowed the system to increase the segmentation accuracy.

# Chapter 5. Texture-based Multi-level Trainable Segmentation

This chapter is concerned proposed with the segmentation of the ROI in ultrasound scans of the ovary in relation to the classification of miscarriage and ovarian tumours, although these two tasks have different objectives as a result of the differences between the GS in pregnancy and a tumour mass. Regarding to the challenges that we have faced in chapter 4, we shall develop a machine learning-based scheme that is referred to as a trainable segmentation method, and highlight its advantages over the method proposed in Chapter 4.

The concept of trainable detection and segmentation has been used for various different applications i.e., content-based image retrieval, face detection, video surveillance, traffic control systems and medical imaging. The widespread use of this concept motived us to use it to enhance the segmentation and address the limitations that we have faced with the solution proposed in chapter 4 (section 4-4). These limitations are: 1) the proposed threshold-based solution does not locate small ROIs, especially in cases where the ROI has a similar texture to the background; 2) does not catch the poor border; and 3) filtering out non-ROI based on geometry features is ineffective as a solution for irregularly shaped objects as well as small objects.

This chapter presents:

1) An effective multi-level segmentation of GS and YS using texture-based trainable models for pregnancy assessment. This represents a new approach to automatic segmentation of the GS and YS from static ultrasound images followed by estimating the MSD of the GS that forms part of the diagnostic criteria for a miscarriage, where an enlarged YS can also indicate pathology. Our approach has an automatic validity check embedded that excludes the images where the GS cannot be successfully segmented. In practice, this means that certain images still need the intervention of the gynaecologist.

2) A trainable segmentation and watershed transform for identifying unilocular and multilocular cysts from ultrasound images of ovarian tumours. A new approach that automatically segments the ovarian masses and cysts from a static B-mode image was used. Initially, the method uses a trained neural network classifier to accurately identify the position of the masses and cysts. After that, the borders of the masses can be estimated using watershed transform.

The rest of the chapter is organised as follows: Section 5.1 sets the scene by briefly describing the problem statements. In Section 5.2, we introduce our new approach to identify the pregnancy stage and extract the GS and YS, followed by MSD measurements. Section 5-3 introduces a new method to segment ovarian masses and cysts from a static B-mode image. Finally, Section 5.4 summarizes the chapter.

## 5.1 Problem Statement

### 5.1.1 GS Images

In order to extract and measure the MSD of the GS and YS for pregnancy cases automatically, it is necessary to first segment the ROI from a B-mode ultrasound image. However, the technical challenges that we have faced with the algorithm proposed in chapter 3 are: 1) artefacts and noise within the image such as speckling, attenuation, signal dropout and shadows can make the task extremely complicated (See Figure 5-1). We solved this problem in chapter 3 by using the wavelet transform to highlight the ROI, but we found that when we have a small GS, or the GS has a texture similar to the background, the wavelet will make the GS disappear rendering the segmentation task impossible. Figure 5-2-A shows example of a difficult case. We applied the threshold after pre-processing (see Figure 5-2-B) and can clearly see that the ROI has been missed. In addition, we have applied the threshold directly to the image without pre-processing (see Figure 5-2-C). The true positive has been located, but there are a number of false positives produced in addition to the ROI. The trainable model (see Figure 5-2-D) improved the result by reducing the number of false positives, which had a positive effect on filtering out non-ROI results. The trainable segmentation stage was followed by the filtering out of non-ROIs to locate the true positive object (ROI). 2) Inhomogeneity of the ROI, which means that it has different textures, and the ROI may be confused with its surrounding areas, causing a problem known as "under-segmentation", i.e., parts of the ROI are not detected. Applying the traditional RG will result in a problem related to the threshold parameter that presents the difference between the pixel and its neighbour. In chapter 4, we will address this problem in detail. 3) The irregularity of the ROI makes the filtering of non-ROI objects based on geometry features extremely difficult. 4) Poor borders (or lack of a clear border) of the ROI and similarity between its texture and that of the background will result in "over-segmentation", where irrelevant parts of the surrounding areas are taken as parts of the ROI. Due to reasons mentioned previously, more than one ROI, i.e., false positive ROI(s), may often be found, causing difficulties in determining which the true positive is. Figure 5-1 illustrates the above problems.

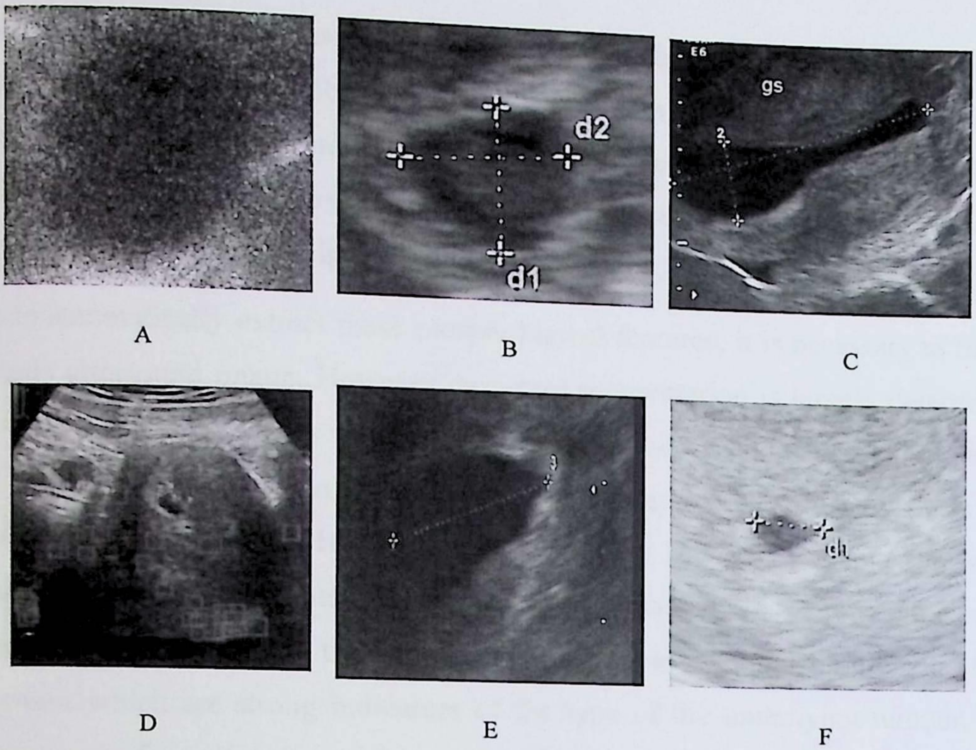



*Figure 5-1: Examples of challenges for correct segmentation of the GS: A) image quality, B) inhomogeneous ROI, C) irregular ROI, D) false positive problem, E) poor border and F) Similarity between the texture of the ROI and the background.*

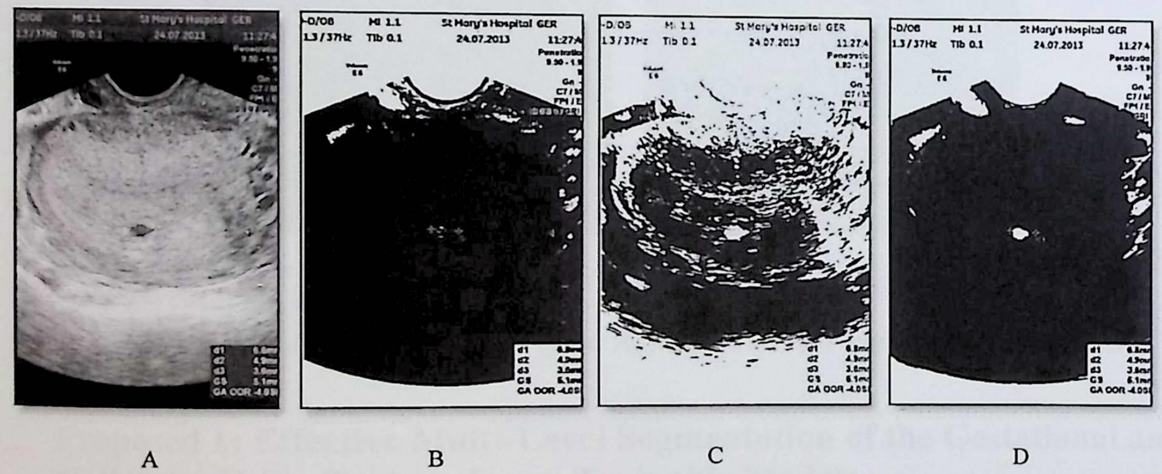
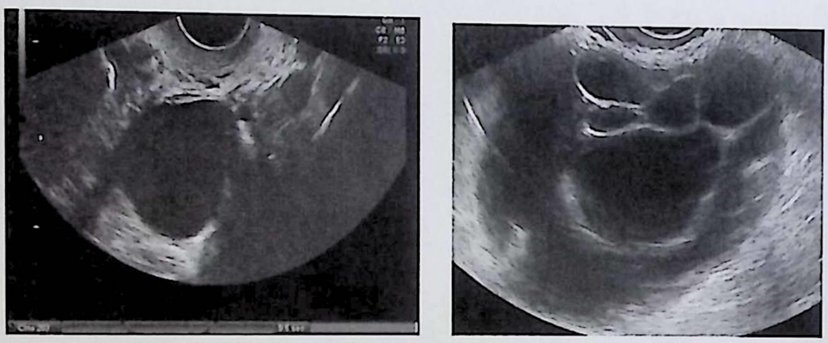



*Figure 5-2: An example of a GS image with small ROI: A) original image, B) image binarization based on a threshold (with pre-processing), C) image binarization based on a threshold (without pre-processing) and D) image binarization based on a trainable model*

### 5.1.2 Ovarian Tumour Images

In order to characterise benign or malignant masses and distinguish tumour types, ultrasound imaging has become the most widely used method in this regard due to its useful imaging characteristics, which can be grouped into two categories. The first of these is the morphological features that are detailed on a B-mode image. These include unilocular and multilocular cysts, fluid, lesion diameter, internal wall structure, papillary projections and acoustic shadows. The second category is the combined use of Doppler images to gain blood flow information. The above enable the clinician to determine the seriousness of an imaged mass (Sayasneh, et al. 2015) (Sayasneh, et al. 2016). Identification of unilocular and multilocular cysts is one of the features affected that can help to identify tumour type.

In order to automatically extract these morphological features, it is necessary to first segment the B-mode ultrasound image. However, accurate segmentation is largely determined by the quality of the image as artefacts which, as mentioned above, make the task extremely complicated. These quality impacts can lead to boundaries being missed, and acquisition orientation of low contrast between the ROIs (Noble and Boukerroui 2006). This chapter presents a novel approach for automatic segmentation of ovarian cysts from static B-mode ultrasound images that allows the number cysts in the image to be counted to distinguish unilocular from multilocular cases, which are strong indicators of the type of the underlying tumour. Figure 5-3 shows two cases of ovarian ultrasound images: A) unilocular case and B) multilocular case.



A B

*Figure 5-3: Unilocular and multilocular cysts: A) unilocular case and B) multilocular case*

## 5.2 Proposed 1: Effective Multi-Level Segmentation of the Gestational and Yolk Sacs Using Texture-based Trainable Models

The mean sac diameter (MSD) of the gestation sac (GS) (see section 2-2-1) forms part of the diagnostic criteria for a miscarriage, and an enlarged yolk sac (YS) can also indicate pathology. This chapter presents a new approach to automatic segmentation of GS and YS from static ultrasound images followed by measurement of the MSD for both YS and GS. To achieve this, the proposed approach involves six steps to segment and measure both the GS and YS. In **Step1,** we locate a set of candidate GSs using a trainable segmentation technique based on the histogram of oriented gradients (HOG) and a neural network (NN) classifier. In **Step2,** we classify each candidate GS into GS or non-GS objects based on the HOG and local binary pattern (LBP) features extracted from each candidate GS is fed into an SVM classifier. To strengthen this step, we injected a reliability check so that instead of locating the sac inaccurately, our method rejects images that cannot be reliably segmented. The output of the previous step is the approximate location of the GS. In **Step3**, the approximate location of the GS is used as a seed mask for a region growing method to obtain and measure the entire GS. **Step4** then automatically identifies the stage of pregnancy based on histogram analysis of the content of the segmented GS. If the pregnancy test indicates the existence of the YS, **Step5** is used to automatically detect and segment the YS. Finally, the size of the YS is automatically estimated in **Step6**.

The main contributions of this proposal are, first, proposing a multi-level trainable segmentation for measuring GS and YS from ultrasound images using texture-based models. Second, filtering out non-sac objects using an object detection model based on the texture feature and SVM classifier. Finally, identifying the stage of pregnancy based on the histogram analysis.

Figure 5-4 provides an overview of the system for the automatic identification of the stage of pregnancy as well as detecting and measuring the GS and YS from a static B-mode image. Each step of the framework will be explained in detail in the following sub-sections.

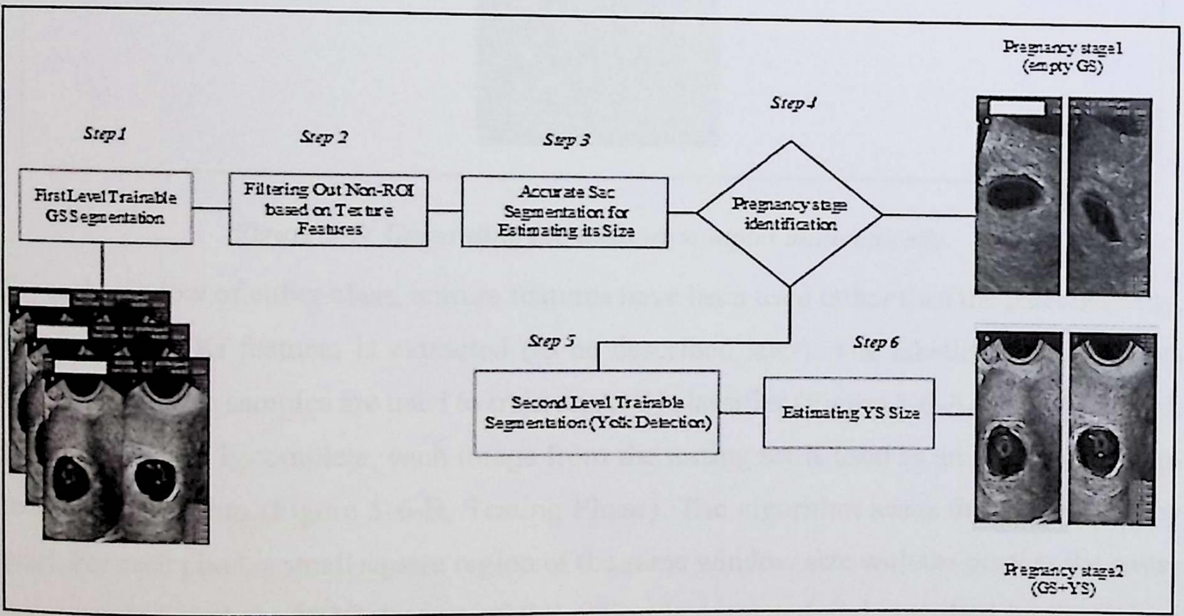



*Figure 5-4: General framework for automatic segmentation*

### 5.2.1 Step1: First Level Trainable GS Segmentation

The aim of the first level segmentation is to isolate the GS from the rest of the image. The process starts with selecting a certain number of images for training purposes. From each training image, a number of samples, i.e., a small window of square regions of a certain size (e.g., 5x5), are taken from inside (Class 1) and outside (Class 2) the GS. To achieve this aim, a semi-automatic tool has implemented to generate the samples. This process starts by manually detecting the border of the ROI. Then, based on the border, we will generate a mask that represents the ROI with 1 pixel and the background with 0 pixels. Based on the mask, a number of blocks are selected randomly from inside and outside the ROI (see Figure 5-5).

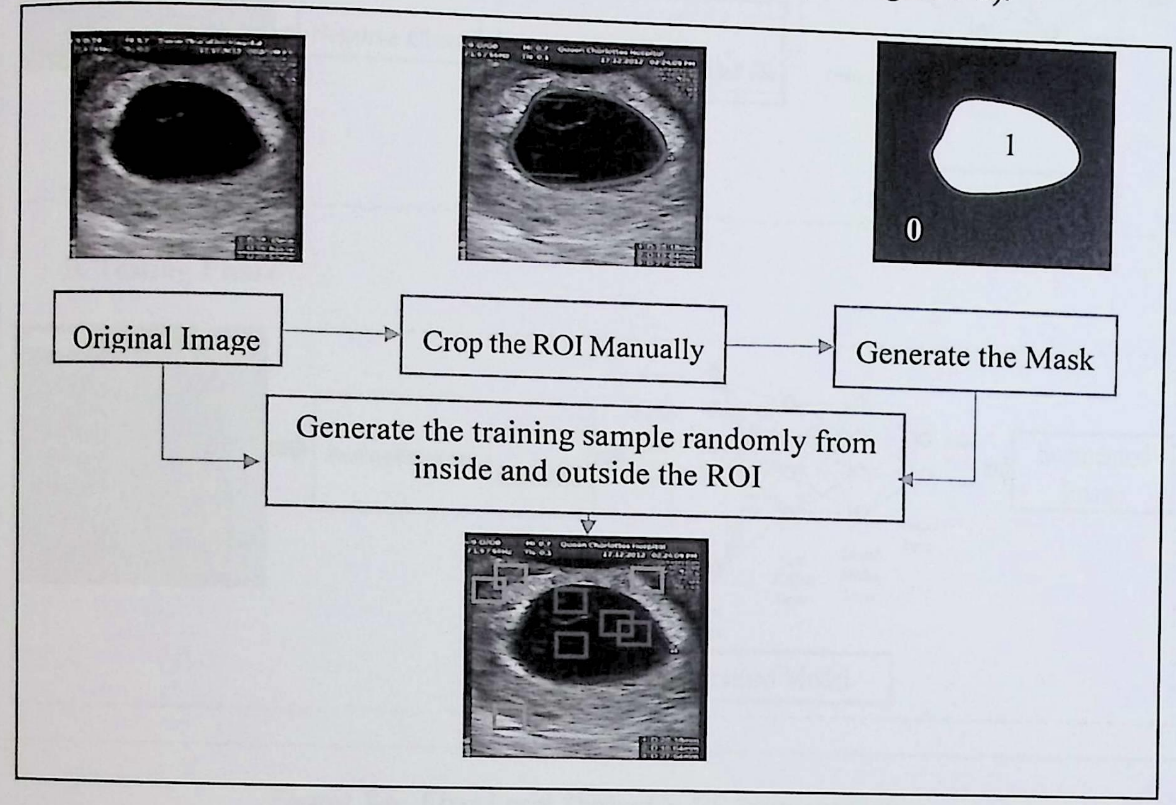



*Figure 5-5: Generating the training samples automatically.*

For each window of either class, texture features have been used rather than the pixel intensity. A set of the HOG features is extracted (to be described later). The labelled feature vectors collected from the samples are used to train an ANN classifier (Figure 5-6-A, Training Phase). Once the training is complete, each image from the testing set is used as an input for the segmentation algorithm (Figure 5-6-B, Testing Phase). The algorithm scans the image pixel by pixel. For each pixel, a small square region of the same window size with the pixel as the centre is constructed, and the HOG feature of the region is extracted and classified by the trained neural network into two classes "inside GS" or "outside GS". Once all the pixels of the image have been labelled, the "inside pixels" region is taken as the segmented GS.

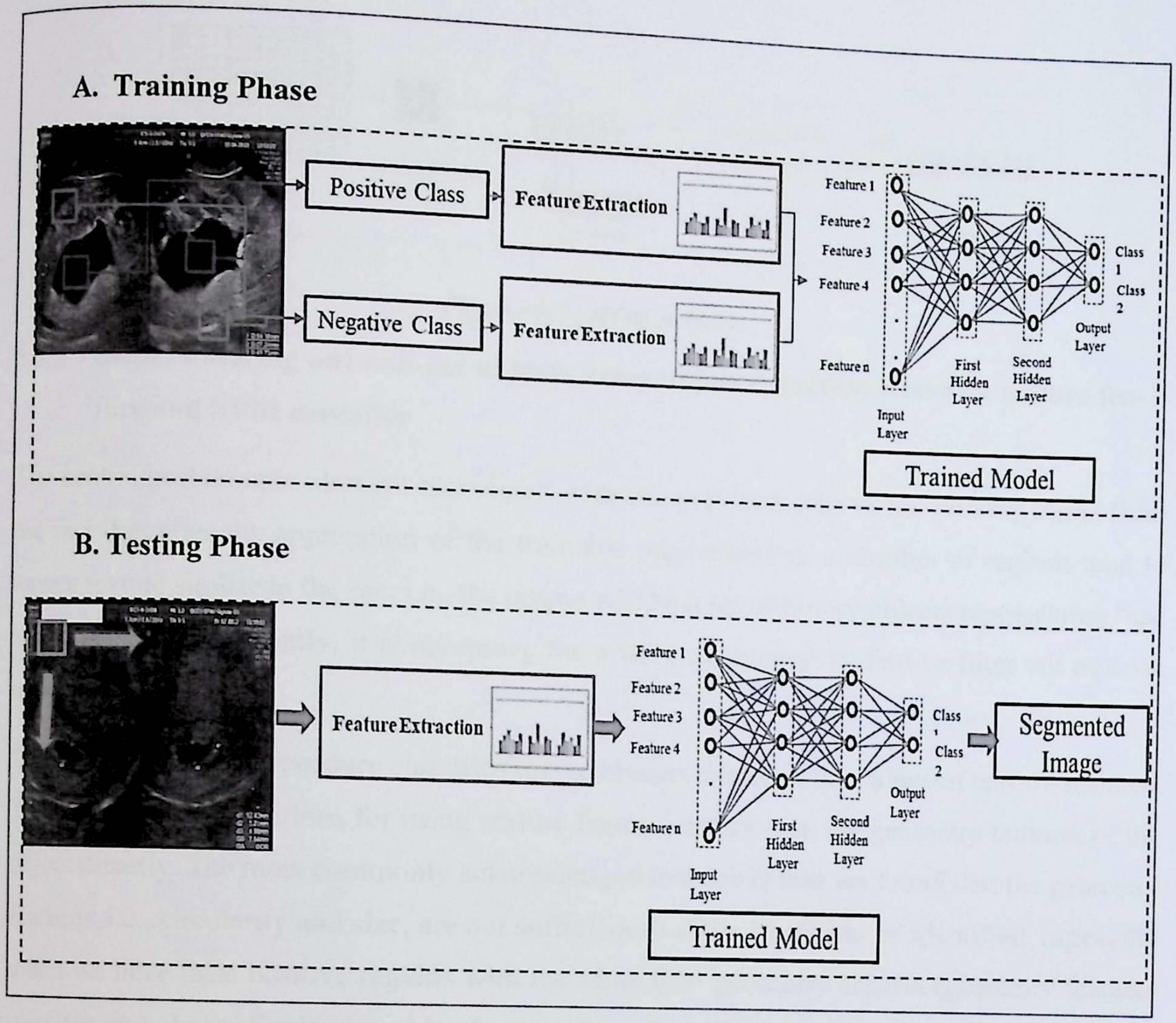



***Figure 5-6: First Level Trainable GS Segmentation***

The histogram of oriented gradients (HOG) is a feature descriptor that facilitates the identification of objects in digital images. HOG was formulated for the quantification of gradient orientation occurrences in localized image sections and has been well documented in (Dalal and Triggs 2005). The process of HOG feature extraction involves taking a window around the pixels called cells. The mask [-1, 0, 1] is used to compute image gradients. In our adaptation of this extraction method (see Figure 5-7) for orientation binning, we directly used the gradient at each image location for the corresponding orientations. The orientation cells are chosen in the range of 0-180° with 9 bins. For better invariance to illumination, shadowing, etc., contrast-normalization of the local histogram is applied.

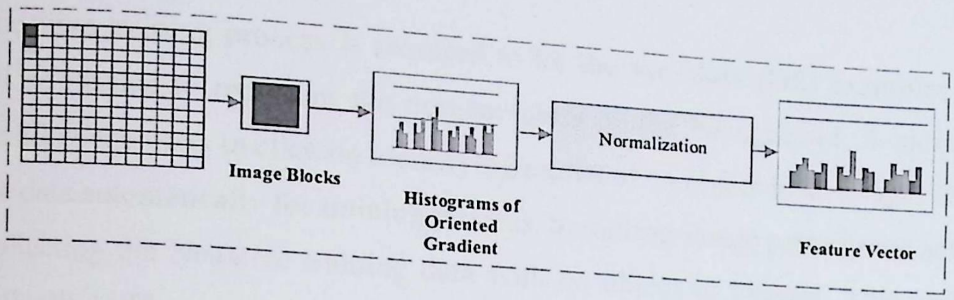



*Figure 5-7: HOG process*

### 5.2.2 Step2: Filtering out non-sac objects using object detection based on texture feature and SVM classifier

The central problematic element associated with the previous step of the process stems from the fact that after the application of the trainable segmentation, a number of regions tend to have a texture similar to the sac, i.e., the output will be a set of binary objects representing "sac candidates". Consequently, it is necessary for a fully automated system to filter out non-sac objects.

Our object detection procedure classifies image objects based on the values of texture features. There are many motivations for using texture features rather than the geometry features of the object directly. The most commonly acknowledged reasons is that we found that the geometry features, i.e., circularity and size, are not sufficient to allow the GS to be identified, especially when we have false positive regions with the same ROI geometry criteria (geometry features overlapping). Accordingly, our objective here is to find a powerful texture feature with which to reduce the overlap between the false positive and true positive classes. These features will support the classifier model to identify the correct ROI. In this model, we have used two features, first, the LBP, which maintains robustness to the monotonic grey-scale changes caused, for example, by variations in illumination. Second, as mentioned earlier, HOG is deployed, which tries to capture the shapes of structures in the region by recording gradient information. One of the advantages of these two features (LBP and HOG) is the production of fixed-length dimension feature vectors for different size boxes due to the number of bins for the uniform LBP and number of angles for the HOG. The results presented in Section 5.2.7 confirm the effectiveness of this fusion (LBP+HOG).

#### A. Training Phase

To understand the organization or establishment of sac-class and non-sac-class pattern, various images of an object will be used. The system input will involve an image consisting of boxes obtained from the GS. During the training phase, data reflecting the GS are needed. This means that a true positive training process is required to be the sac-class (GS) examples and false positive will be needed to represent the non-sac-class or the background. A simple tool is developed to allow the users to click on identifying marks of an object in an image. This assists in cutting the data automatically for training process. Sampling image patterns are achieved by randomly collecting the Negative training data with no object of interest. This tool was designed specifically to filter out non-ROI because, in this stage, we want to include the edge of the GS, where the texture of this edge will help to filter out the FP cases during the testing stage. The HOG and LBP operator is used to extract the texture feature vectors for both positive and negative classes. These features will be fed to the SVM classifier with their labels to build a trained model structure. The justification for choosing the SVM at this stage will be discussed in the following subsections.

**B. Testing Phase**

The output of this First Level Trainable GS Segmentation is a binary image which includes the GS and some non-GS objects. A rectangle is drawn around the objects, first, to get the edge of the GS which we strongly believe contains good information for identifying the GS (positive class) from the negative class and, second, the calculation of the HOG and LBP features has been simplified. The objective of this method is to detect the correct ROI. Figure 5-8 shows how the system automatically draws a rectangle around the candidate GS. To get a rectangle around the binary object according to the XY-axes, the proposed system needs to find the min/max of all the $x$ and $y$ coordinates. Then, to cover the edge of the region, for the proposed solution N pixels around the object are included. Based on our observation, N=5 seems optimal to cover the edge of the GS. Figure 5-9 shows the proposed detector.

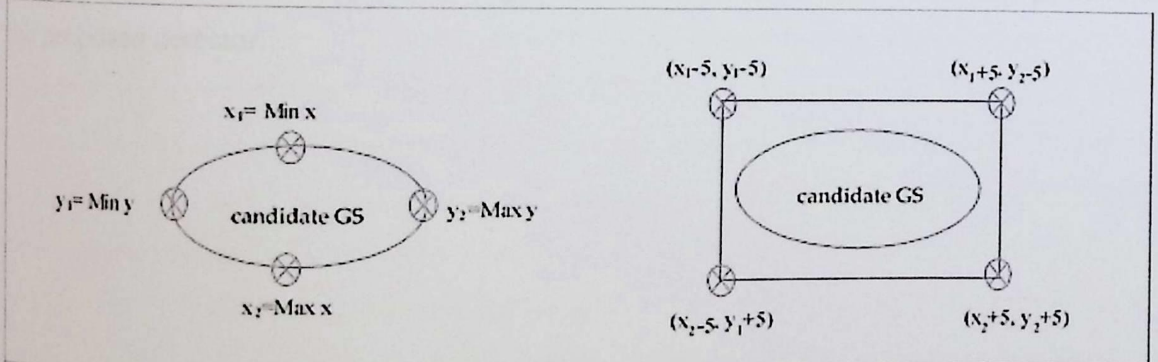



*Figure 5-8: Drawing a rectangle around the candidate GS*

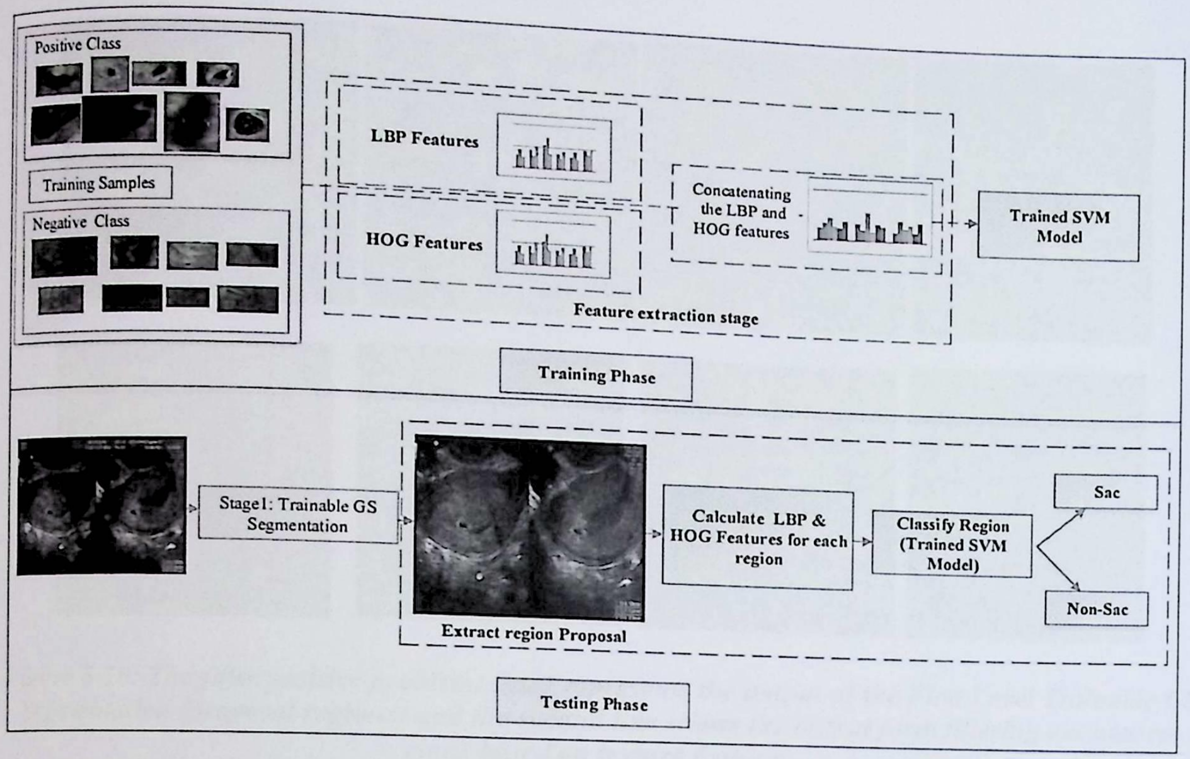



*Figure 5-9: Filtering out non-regions based on texture features*

As mentioned previously, we have used the SVM classifier to identify the sac, and from which there are three possible scenarios as the outcome of this model. First, we will get only one sac. Second, we end up with none i.e., no sac can be reliably detected, that is, it can tell us there is no true positive case when the First Level Trainable GS Segmentation did not extract the GS. In this case, our method rejects the image as it cannot be reliably segmented. Third, we will end up with more than one candidate sac. In this case, the SVM score used to rank the candidate sacs and the candidate with the highest score represents the correct GS. Figure 5-10 shows the proposal regions from the first level segmentation and the selection of the true positive using the proposed detector.

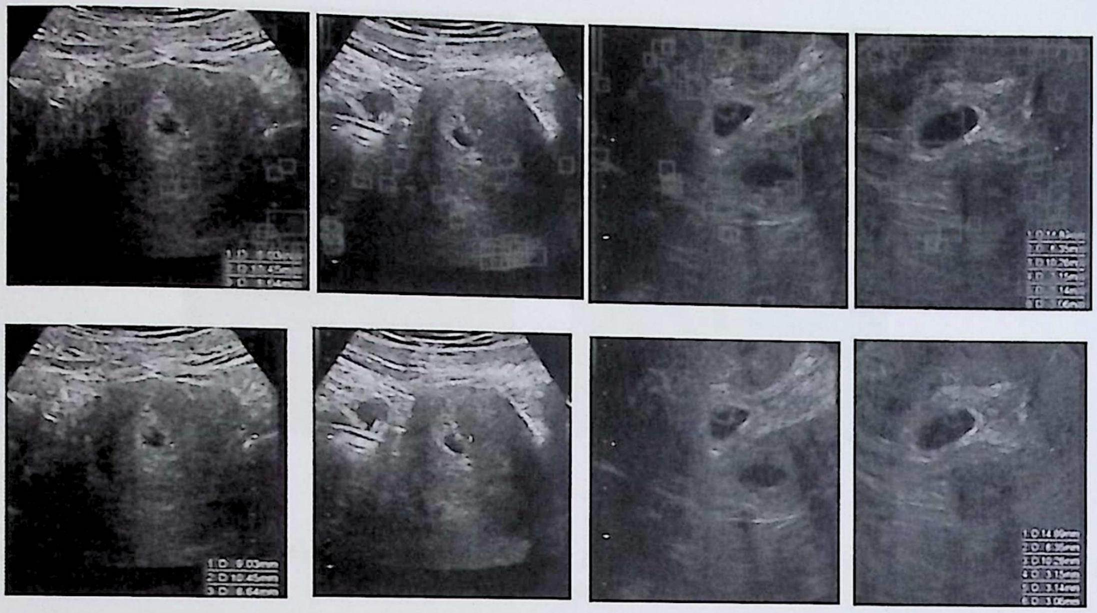


*Figure 5-10: The false positive problem: line1 represents the output of the First Level Trainable GS Segmentation (proposal regions) and the second line shows the output form filtering out non-regions based on texture features.*

Finally, three questions related to the Step2 can be raised.

- **Why the SVM model as a classifier in Step2?**

In Step2 we trained the SVM with two classes (sac and non-sac). Each image has only two GS and, due to the limited number of images, we do not have enough sacs to train the ANN. In addition, the SVM score has helped to select the correct GS. By contrast, in Step1 we have used small window to train the classifier which helped gain a good number of samples from inside the GS. Based on this strategy, we gained enough data to train the ANN in Step1.

- **Why the HOG in addition to the LBP?**

The HOG features method is based on a gradient image method. Therefore, it captures the shape of structures in the region by recording information about the edge of the gradient image. Figure 5-11 shows the gradient image for a case of a GS as well as another object (false positive object) with a similar gradient structure. Therefore, using the HOG alone is not sufficient to locate the correct GS. Figure 5-12 shows the LBP of the correct GS and the false positive. We can see that the LBP can capture the monotonic greyscale changes as well as the borders that can help to identify the correct GS. Based on our investigation and the results discussed in section 5.2.7 we can confirm that using the LBP and HOG will help to capture different information to reduce the overlap between the true positive regions (GS) and the false positive regions.

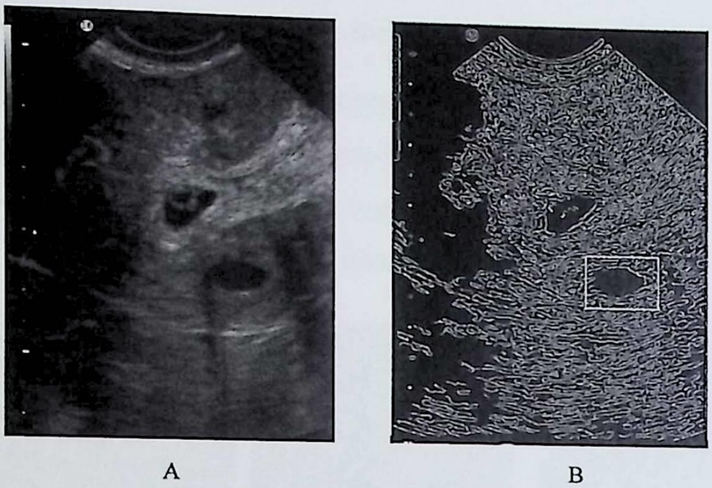


Figure 5-11: Illustration of the gradient image as a part of the HOG: A) original image and B) gradient image

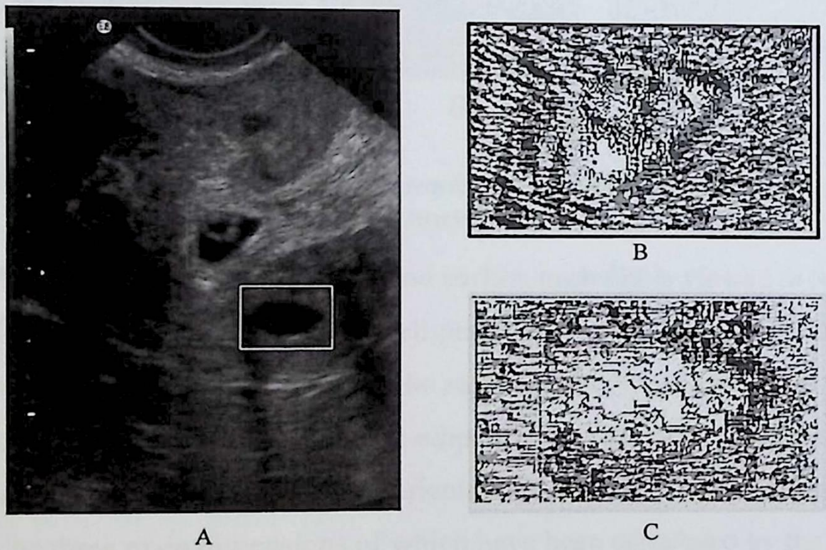


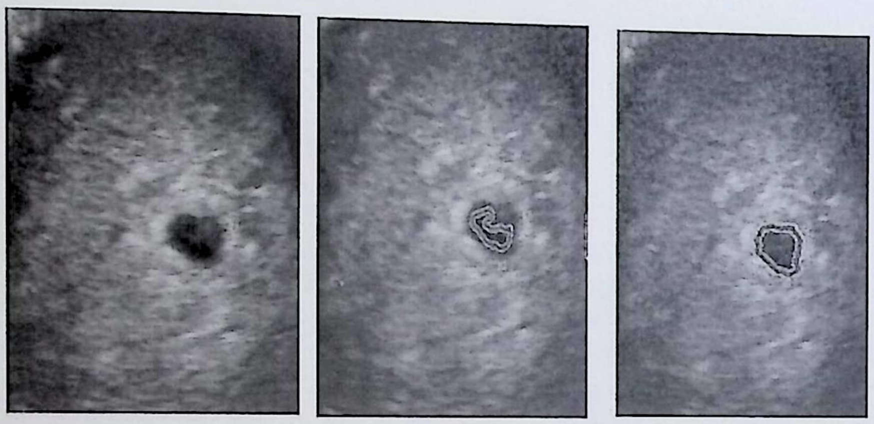
Figure 5-12: LBP for candidate: A) original image, B) LBP for right GS and C) LBP for the false positive object

- **Why LBP and HOG in step 2 and in step1 just HOG?**

The LBP with HOG was not used in step1 for two reasons. First, the output of the LBP is a histogram with 255 or 58 bins. Representing a window with size 5*5 or 7*7 with this number of features means the majority will be zeros. Second, as we can see from Figure 5-11-B that the HOG based on the gradient image can distinguish the area inside the ROI from the background.

### 5.2.3 Step3: Accurate Sac Segmentation for Estimating its Size

The output of the previous step gives an approximate location for the GS. In this step, this location is used as a seed mask for other segmentation techniques to obtain the complete GS. Region growing is used in this step as a post-segmentation technique to extract the sac with the correct boundary; Figure 5-13 shows the results of this use.



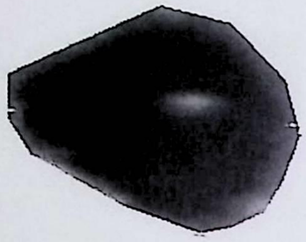
A B C

*Figure 5-13: Shows the result of the region growing: A) original image, B) seed mask and C) Region growing result*

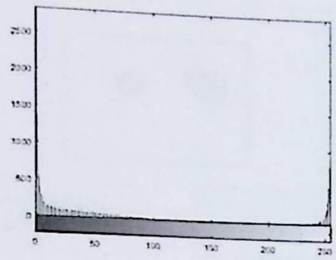
To measure the MSD for the GS, as explained earlier, each GS is viewed in two perpendicular planes. The GS usually seems to be more elliptical in its early stages, and hence our system attempts to locate the best-fitted ellipse for the segmented GS for each plane. The region props function in MATLAB was utilised to fit an ellipse to the GS, which returns four parameters: the major axes, minor axes, centroid and orientation. The GS presumably has an ellipsoidal shape in 3D, the three main dimensions of which have been calculated by the minor and major dimensions from the sagittal plane and the major dimension from the transverse plane. The average of the three dimensions is taken as the MSD.

### 5.2.4 Step4: Histogram Analysis to Identify the Pregnancy Stage

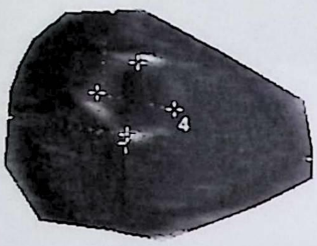
This stage aims to establish if the GS is empty or not, i.e., whether it contains a YS. To do this, first, the binary image resulting from the GS segmentation step is used as a mask to locate the pixels inside the GS from the original image. Histogram analysis is applied to these pixels to determine whether the YS is present within the GS. Figure 5-14 shows the histogram of the GS and GS+YS, where the difference between them is immediately obvious. Figure 5-15 shows the process used to identify the empty GS from the GS that has a YS.



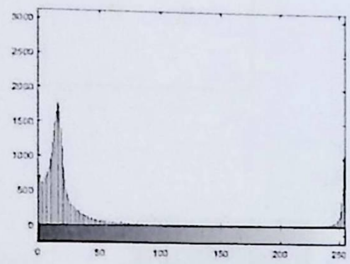
A) Empty GS



B) histogram of the empty GS



C) GS + YS



D) histogram of the empty GS+YS

***Figure 5-14: Histograms for both GS and the GS+YS: A) empty GS, B) histogram of the empty GS, C) GS with the YS and D) histogram of the GS and YS***

Consider that the grey level histogram matches with a GS, f $(x, y)$, constituted of light objects (the YS's border) superimposed on a dark backdrop (the GS). Furthermore, the configuration is organized such that the pixels of the object and backdrop display grayscale levels are categorized into a pair of dominant modes. If this is the case, the immediately discernible method by which the border of the YS can be estimated from the GS involves the selection of a threshold (T) that facilitates the separation of these modes. Consequently, a pixel $(x, y)$ according to which f $(x, y) > \mathrm{T}$ (based on bin frequency) can be designated as a YS border.

Under certain circumstances, when the histogram frequencies of the greyscale values are high within the GS, it does not necessarily mean there is a YS; the frequency simply serves to denote small noise objects. Consequently, for each bin with frequency greater than T, pixels of the intensity represented by the bins are located in the GS image (see Figure 5-15). A post-processing check is conducted to determine if the pixels form an object. If not, the GS is considered to be empty (stage1); otherwise, the GS contains a YS (stage2).

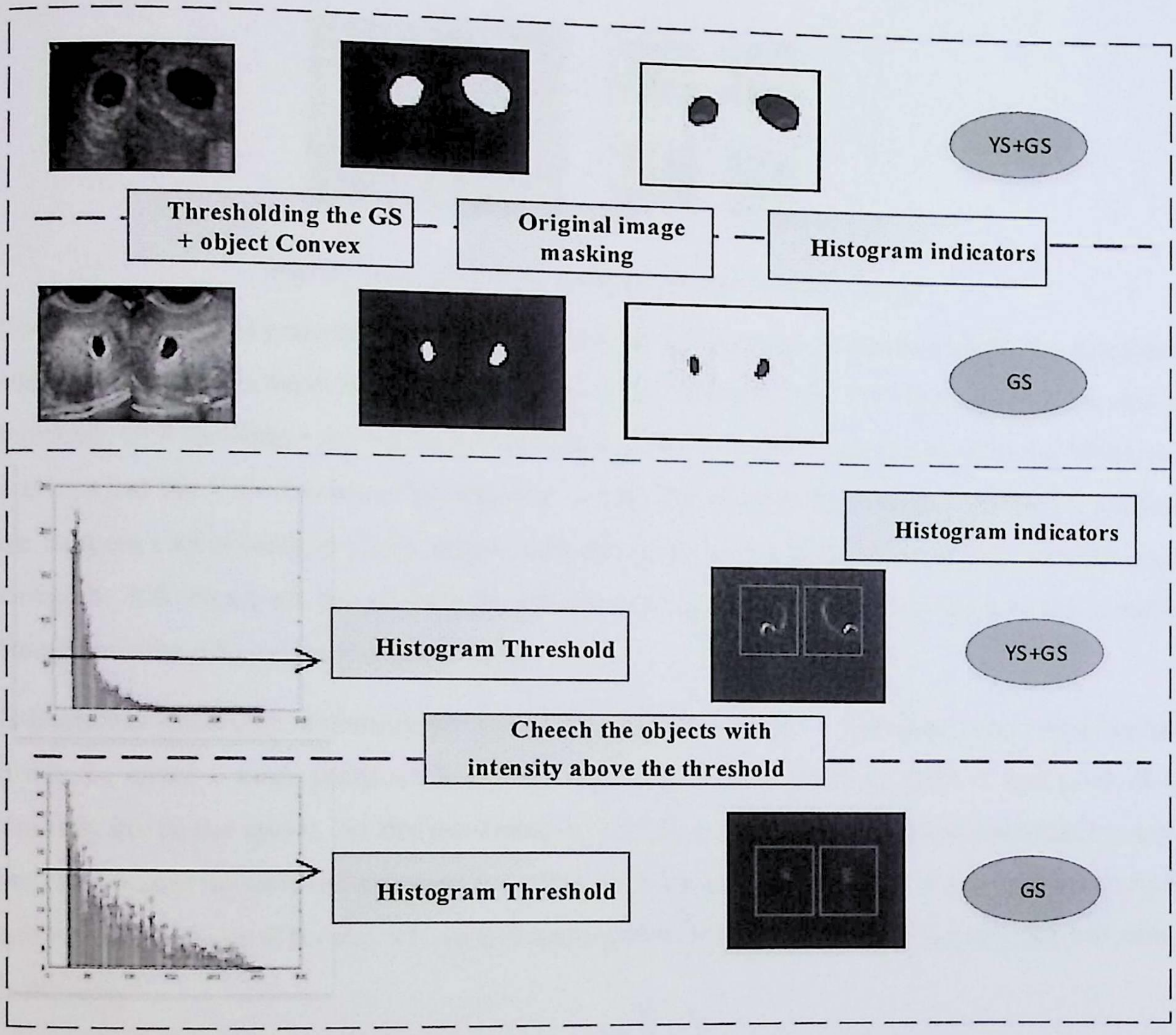



*Figure 5-15: Histogram analysis to identify stages of pregnancy*

### 5.2.5 Step5: Second Level Trainable Segmentation: Estimating the YS Border

Once the histogram analysis of the area inside the GS establishes the existence of the YS, a second level trainable segmentation is employed at this step to estimate the YS border. The main challenge here is that in certain scenarios this border is vague and unclear, which means that techniques based on the pixel intensity thresholding cannot work satisfactorily. The machine learning-based approach provides more promise. A pixel feature (in accordance with the pixel neighbourhood) is used to identify the border pixels. The training phase is implemented by selecting a window of size 5×5 for each yolk boundary and non-boundary in the image (see Figure 5-16). The HOG operator is used to extract the texture feature for both border and non-border YS. The testing phase will process only the pixels that are inside the GS sac by using the mask generated by the previous segmentation steps. The output of this step is a binary GS including the GS in white and the YS border in black.

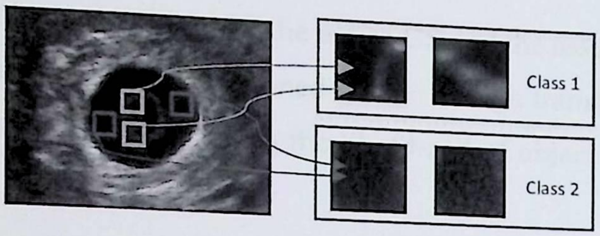



*Figure 5-16: Selecting samples for the learning phase*

Once the trainable segmentation has estimated the YS border, a convex hull (to be described later) is employed to facilitate the delineation of the YS from the GS. The rationale for this is threefold: (i) it facilitates the maintenance of the entire GS and the assessment of the MSD; (ii) it allows just the YS's border to be retained for the Hough transform stage; and (iii) it reduces the frequency of objects in conjunction with the likelihood of object circularity, thereby capturing the YS. Next, all the objects inside the convex hull are used as input to the circular Hough transform to detect the best circle.

In Euclidean geometry, a convex set denotes a space for which – regarding each pair of points inside the space – each point on a straight-line segment connecting point A and point B is similarly inside the space. As demonstrated in Figure 5-17, a filled-in circle constitutes a convex set whereas an unfilled crescent (or filled-in crescent) does not. In essence, the set (S) is referred to as convex if for any x1, x2 ∈ S and any θ with $0 \leq \theta \leq 1$. Consequently, the following is the case:

$$\theta x1 + (1 - \theta)\, x2 \in C. \qquad (5\text{-}1)$$

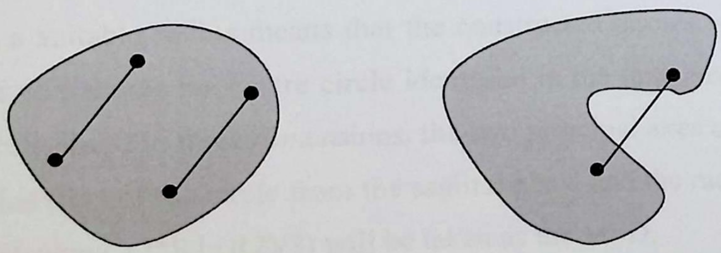



*Figure 5-17: simple convex and non-convex sets. Left. The circle, which includes its boundary (shown darker), is convex. Middle. Right. The kidney shaped set is not convex.*

A convex hull is the minimal possible convex region with the capacity to encompass a predetermined set of points. Calculating a finite set convex hull is a central practice in numerous disciplines, including image processing, and it occupies an important place in the present research insofar as it facilitates the delineation of the YS from the GS. The rationale for this is twofold: (i) it facilitates the maintenance of the entire GS and the assessment of the MSD; and (ii) it allows just the YS's border to be retained for the Hough transform stage. This reduces the frequency of objects in conjunction with the likelihood of object circularity, thereby capturing the YS.

### 5.2.6 Step6: Locating the YS and Measuring its Size

This step aims to identify the YS using some edge generated from the Second Level Trainable Segmentation. The circle Hough transform is used to estimate the edge of the YS. When several points residing on the perimeter of a circle have been identified, this method constitutes an effective way to identify the limits of that circle. The sole point of distinction between this method and the linear Hough transform is that a circle's point on the $(x, y)$ plane is translated into three-dimensional limits, and the expression for the circle is as follows:

$$r^2 = (x-a)^2 + (y-b)^2 \qquad (5\text{-}2)$$

where $a$ and $b$ represent the circle's centre coordinates; $x$ and $y$ represent the perimeter coordinates; $r$ is the radius. Based on the assumption that a circle exists in ($x$ and $y$) space, when the image coordinates, and the YS's perimeter points, have been acquired, a circle in the ($a$ and $b$) parameter space corresponds to every point on its perimeter. The ($a$ and $b$) space constitutes a circle accumulation of the input image for a specified radius ($r$) (see Figure 5-18). The critical piece of information when attempting to identify a circle is the radius, because it determines the circle dimensions in the ($a$ and $b$) space. In the case where a circle constructed in the ($a$ and $b$) space is not identical in size to the initial circle's radius, the former will not contact at a single location. Hence, a suitable radius means that the constructed circles will contact at a single location, and this constitutes the centre circle identified in the image coordinates. Assuming the YS has a circle shape in three dimensions, the two principal axes of the circle can be estimated by the radius (R1) of the circle from the sagittal plane and the radius (R2) of the circle from the transverse plane. (2*R1+R2)/3) will be taken as the MSD.

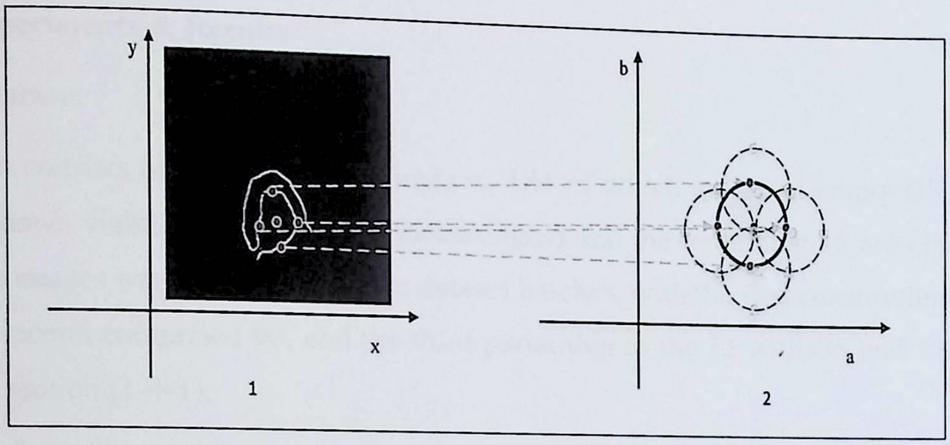



*Figure 5-18: Circle located on the parameter plane x, y, 2) transform the circle to the a and b space*

In the context of the Hough circle transform, the discrete transform parameter space [a b r] constitutes a three-dimensional matrix. Consequently, the detection of the circle in the input image simply involves identifying the greatest frequency of circle intersections. The reader can consult Figure 5-19 for an illustration of the three-dimensional discrete Hough space.

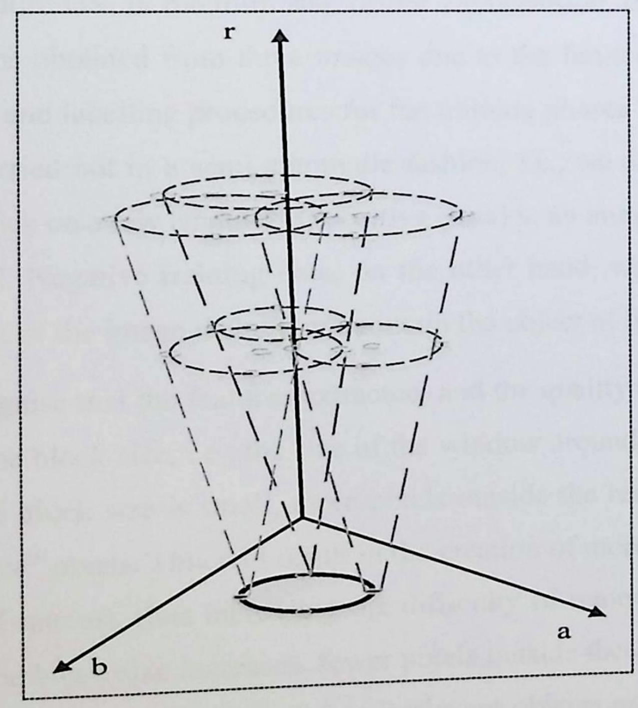



*Figure 5-19: Three-dimensional discrete Hough accumulator space*

The extraction of the YS involves identifying the circles for a certain radius $r_j$ in the parameter space in accordance with every point $(x_i, y_i)$ which belongs to the contour line of a certain circle.

### 5.2.7 Experiments & Results

#### 5.2.7.1 Dataset

The dataset consists of 199 ultrasound images, 184 of which are of an empty GS (157 pregnancy unknown viability PUV and 27 miscarriages) and the remaining 15 are GSs with YSs inside. The images were acquired in three dataset batches, with the first constituting 94 images, whilst the second comprised 90, and the third pertaining to the 15 with GS and YS. For more details see section (1-4-1).

#### 5.2.7.2 Experimental protocol

A two-layer back propagation neural network has been used in each of the three trainable segmentation steps (step1, step2 and step5). In the first step, the network was trained with 500 sample regions obtained from 25 training images. In the second step, we trained the model with the same number of images as in the previous step by cropping regions manually from the background and the entire GS. In the third step of the segmentation, the network was trained with 60 sample regions obtained from three images due to the limited number of images in batch 3. The cropping and labelling procedures for the training phases in the three steps of the segmentation were carried out in a semi-automatic fashion, i.e., we developed a simple tool that allows a user to click on a few landmarks (positive class) in an image to allow it to become automatically cropped. Negative training data, on the other hand, was gathered by random sampling from the part of the image that did not contain the object of interest.

It is important to recognise that the features extracted, and the quality of segmentation are directly influenced by the block size, i.e., the size of the window around the pixel. As shown in Figure 5-20, when the block size is small, more pixels outside the real region of interest are considered to be "inside" pixels. This will result in the creation of more irrelevant objects outside the real object of interest, thus increasing the difficulty of removing those objects later (Figure 5-20-B). As the block size increases, fewer pixels outside the object of interest would be confused as "inside" pixels, and hence, fewer irrelevant objects are created (Figure 5-20-(C, D)). However, when the block size gets too large, the possibility that a block contains pixels of both "inside" and "outside" also increases, resulting in an imprecise classification of the pixel class (Figure 5-20-D)). Hence, to ensure high-quality segmentation, the block size must be determined carefully. Based on our observation, a block size of 7x7 seems optimal due to the decreased the number of false positives and the region growing used as the post-segmentation method which enhances the segmentation border.

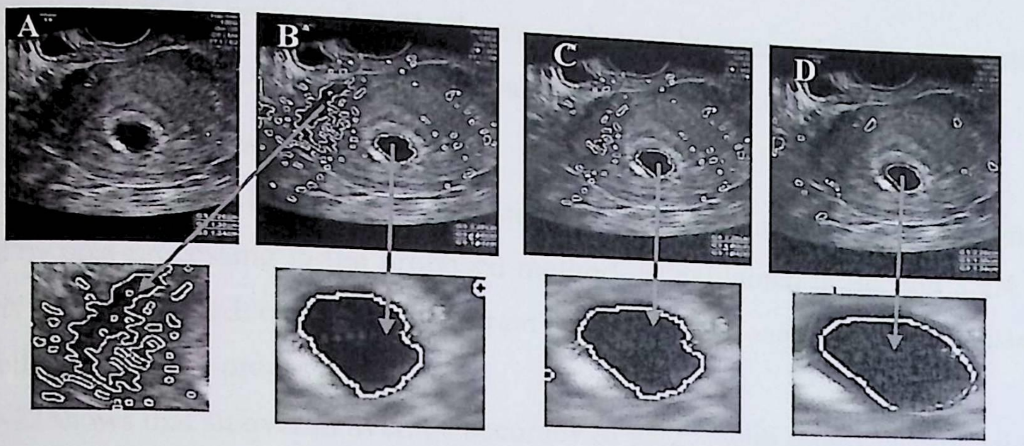



*Figure 5-20: Trainable segmentation result: A) original image, B) segmented image with window size 3x3, C) segmented image with window size 5x5, and D) segmented image with window size 7x7. The white line represents the first step of the segmentation, and the r*

To evaluate the power of the texture feature (HOG or LBP), we have used the Receiver Operating Characteristic (ROC) curve to analyse performance when using the different texture feature. The ROC curve is used to visualise the performance of the classifier, which is commonly deployed in medical image decision making and has recently been increasingly utilised in object detection research. Figure 5-21 shows the performance of the features that have been used to detect the sac in the second segmentation step. The area under the curve (AUC) for the LBP = 0.88 and for the HOG = 80%. When we combine the two features (LBP+HOG), the performance of the SVM is enhanced (AUC = 92%). Based on this evolution we have used the two features to detect the sac (enhancing the true positive rate).

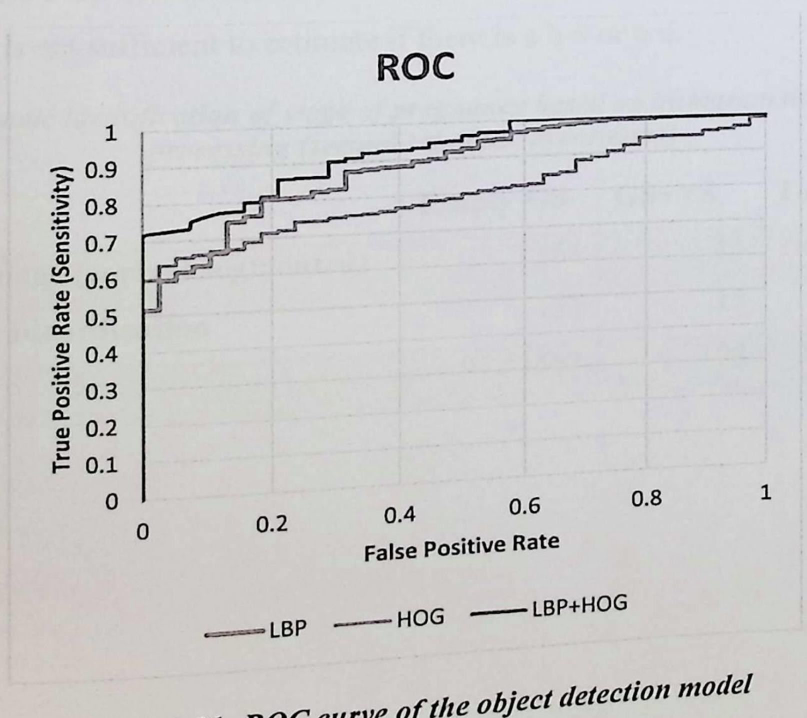



*Figure 5-21: ROC curve of the object detection model*

The effectiveness of the proposed method was evaluated by examining the segmentation precision by comparing the automatic and manual measurement results of the MSD.

#### 5.2.7.3 Pregnancy stages identification

We are interested in the effectiveness of histogram thresholding and histogram thresholding followed by post-processing in the proposed method, as explained in Section 5.2.4. Two experiments were conducted: one using histogram thresholding alone, and the other using thresholding followed by post-processing.

Figure 5-22 shows that an overall average accuracy of 85.2% with stage1 identification (empty GS) of 83.8% and Stage 2 identification (GS &YS) of 100% for can be achieved in the first set of experiments (histogram threshold) (see Table 5-1).

***Table 5-1: Automatic identification of stage of pregnancy based on histogram threshold (first set of the experiment)***

| | **Empty GS** | **GS+YS** | **Total** |
|---|---|---|---|
| **Total Image (correct segmented)** | 161 | 15 | 176 |
| **Correct Identification** | 135 | 15 | 150 |
| | 83.85093 | 100 | 85.22727 |

By contrast, an overall accuracy of 97.7% with stage1 identification of 97.5% and stage2 identification of 100% was achieved in the second experiment (histogram threshold with post-processing) (see Table 5-2). The results show that the second method confirms that the threshold of the histogram is not sufficient to estimate if there is a YS or not.

***Table 5-2: Automatic identification of stage of pregnancy based on histogram threshold with post-processing (second set of the experiment)***

| | **Empty GS** | **GS+YS** | **Total** |
|---|---|---|---|
| **Total Image (correct segmented)** | 161 | 15 | 176 |
| **Correct Identification** | 157 | 15 | 172 |
| | 97.51553 | 100 | 97.72727 |

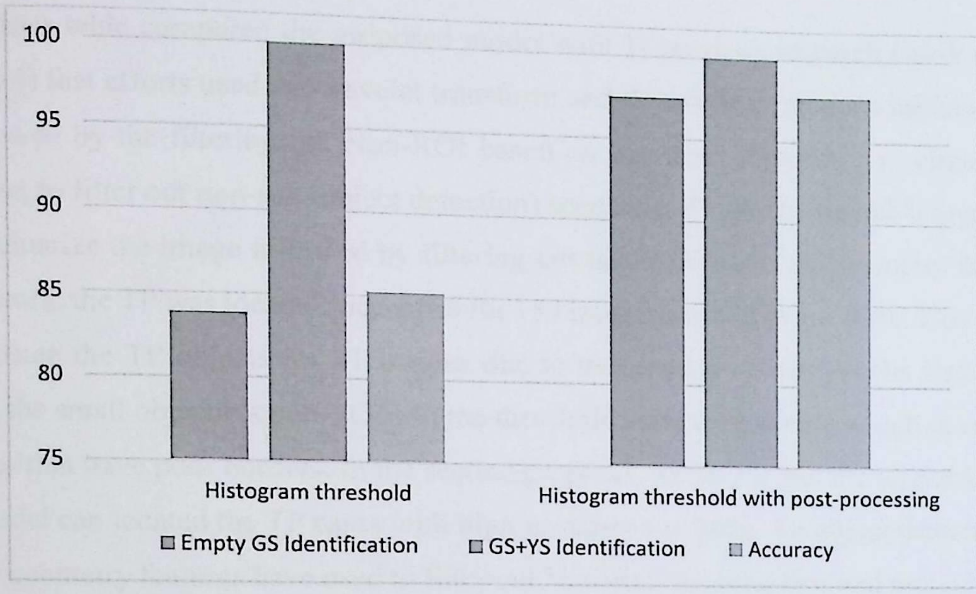



*Figure 5-22: Stage of pregnancy identification*

#### 5.2.7.4 Segmentation Result (Manual vs. Automatic Measurements)

The manual calculation of the proportion of successful GS segmentations was conducted with the human visual system (HVS). Regarding batch 1 and 2, we compared the performance of this algorithm with previous work (Table 5-3).

*Table 5-3: The accuracy of the proposed system compared with previous work*

| | Work reported in section 4-4 | Trainable segmentation with Filtering out non-ROI based on Geometry features | The proposed method |
|---|---|---|---|
| **Located the TP besides the FP** | 151/184 | 177/184 | 177/184 |
| **Accuracy** | 82.06% | 96.19% | 96.19% |
| **Success out of the first step (initial segmentation)** | 137/151 | 153/177 | 161/177 |
| **Accuracy** | 90.7% | 86.44% | 90.9% |
| **Successful segmentation out of the whole dataset** | 137/184 | 153/184 | 161/184 |
| **Accuracy** | 74.4%. | 83.15% | 87.5% |

The previous table compared the proposed model with 1) previous research (work present in section 4-4) that efforts used the wavelet transform and threshold to produce initial segmentation followed by the filtering out Non-ROI based on geometry features, i.e., circularity and object size, to filter out non-sac (object detection) scenarios. 2) use the trainable segmentation stage to binarize the image followed by filtering out non-ROI based on geometry features. In the first work, the TP was located successful for 153 images out of 184 i.e. 82%. This algorithm cannot locate the TP objects for 31 images due to two reason a) the wavelet transform has removed the small objects (small ROI), b) the threshold sometime is not enough to capture the ROI cases that have poor borders. In the second proposal, we found that the trainable segmentation model can located the TP cases with high accuracy i.e. 96%. An object detection model based on geometry features have used to filter out Non-ROI for both first and second proposal. In some cases, the proposed object detection model has failed to identify the correct ROI an thereby resulting in some false positive cases. Therefore, we have used the texture feature in the filtering out stage to identify the TP from the FP objects. Table 5-3 shows clearly the effective of the texture features, where the accuracy increased from 83% to 87%.

Additionally, of the 15 images in the third batch, the GS was segmented accurately in all cases. Furthermore, the YS segmentation was successful for 14 out of the 15 images, i.e., 93.3% of the time. To facilitate the provision of a suitable system of assessment regarding the subsequent phases, each image associated with unsuccessful segmentation of the GS was not included.

Automatic MSD measurements were considered in relation to manual measurements to assess the degree to which the proposed system was effective. Figure 5-23 provides an indication of the performance: Figure 5-23-A illustrates the results of batch 1 and batch 2 (Empty GS); and Figure 5-23-B represents the result for batch 3, where the points inside the red ellipse pertain to the measurement for YS and the other points are the GS measurements. In view of the approximate 45° regression line, it can be concluded that the automatic measurements are remarkably close to the manual ones ( see Table 5-4). It can also seem reasonable to claim that there is no apparent systematic bias to the automatic measurements.

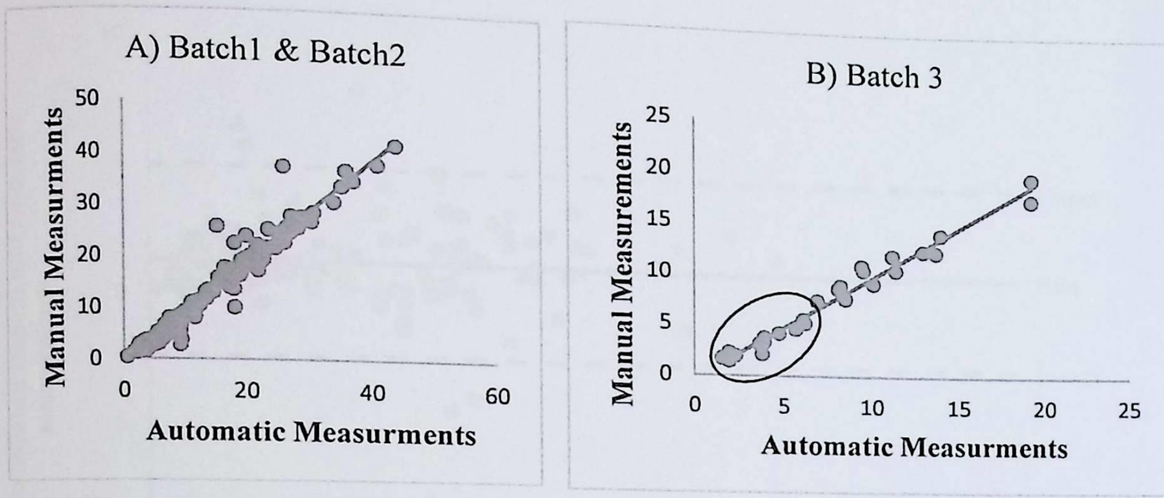



*Figure 5-23: Comparison between manual and automatic MSD measurements for both Empty GS and GS+YS: A) Batch 1&2 and B) Batch 3.*

*Table 5-4: Shows the regression statistics for Batch1, Batch 2 and Batch 3*

| Regression Statistics | | |
|---|---|---|
| | Batch 1 & 2 | Batch 3 |
| **Multiple R** | 0.976877 | 0.987288 |
| **R Square** | 0.954289 | 0.974737 |
| **Adjusted R Square** | 0.954001 | 0.973765 |
| **Standard Error** | 2.003857 | 0.795361 |
| **Observations** | 161 | 28 |

From the $R^2$ value in the scatter plots (Figure 5-23), it is evident that both automatic and manual measurements are highly correlated. This correlation depends on the range of the true quantity in the sample; if this is wide, it will be greater than if it is narrow. The Bland Altman procedure has been used to investigate the real closeness between the manual and automatic measurements. Figure 5-24 illustrates the results for batch 1 and batch 2 (empty GS). Figure 5-25 shows those for batch 3, where the points inside the red ellipse represent the measurement for YS and the other points, the GS measurements. Table 5-5 shows the lower and upper LoA for both Figure 5-24 and Figure 5-25. In addition, the mean square error (MSE) between manual and automatic measurement is calculated, where the MSE for batch 1 and batch 2 is 3.16, whilst that for batch 3 is 1.42 (where the MSE for the GS is 1.68 and for the YS is 1.16).

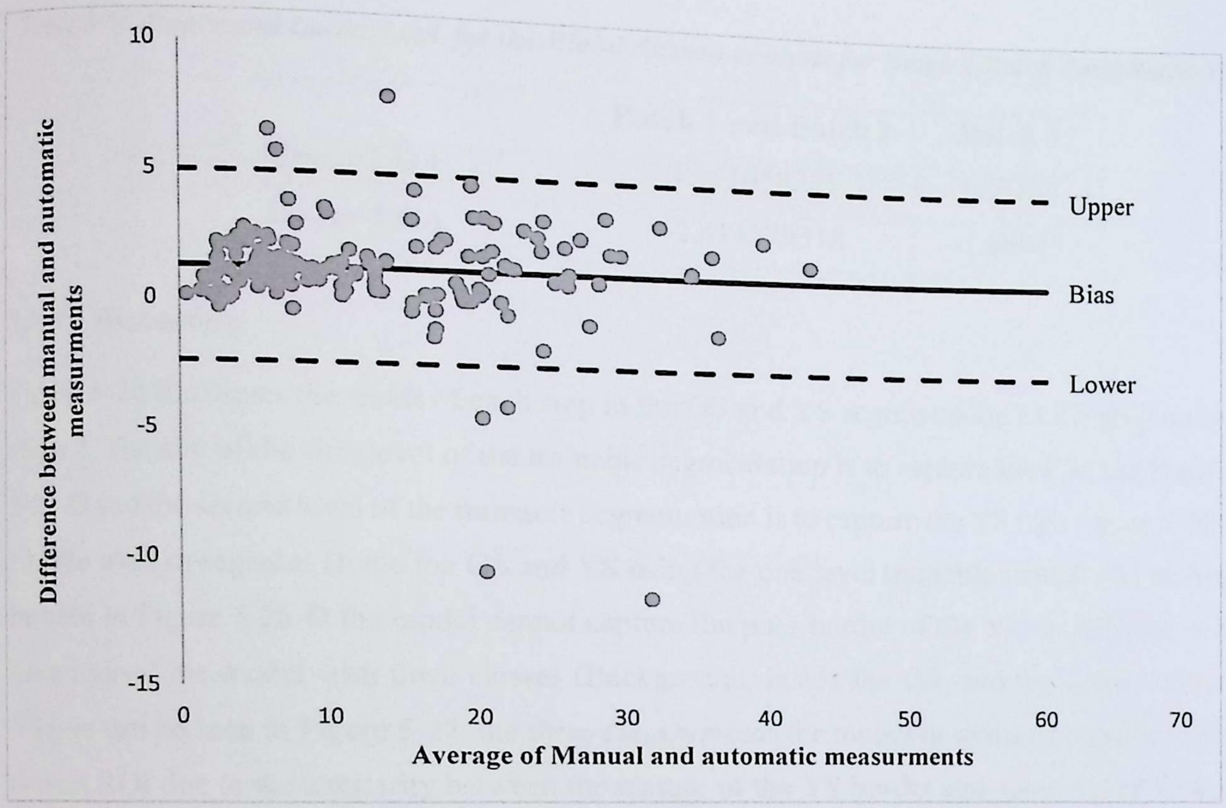



*Figure 5-24: Bland Altman correlation analysis of automatic and manual measurement for batches 1 and 2 (empty GS)*

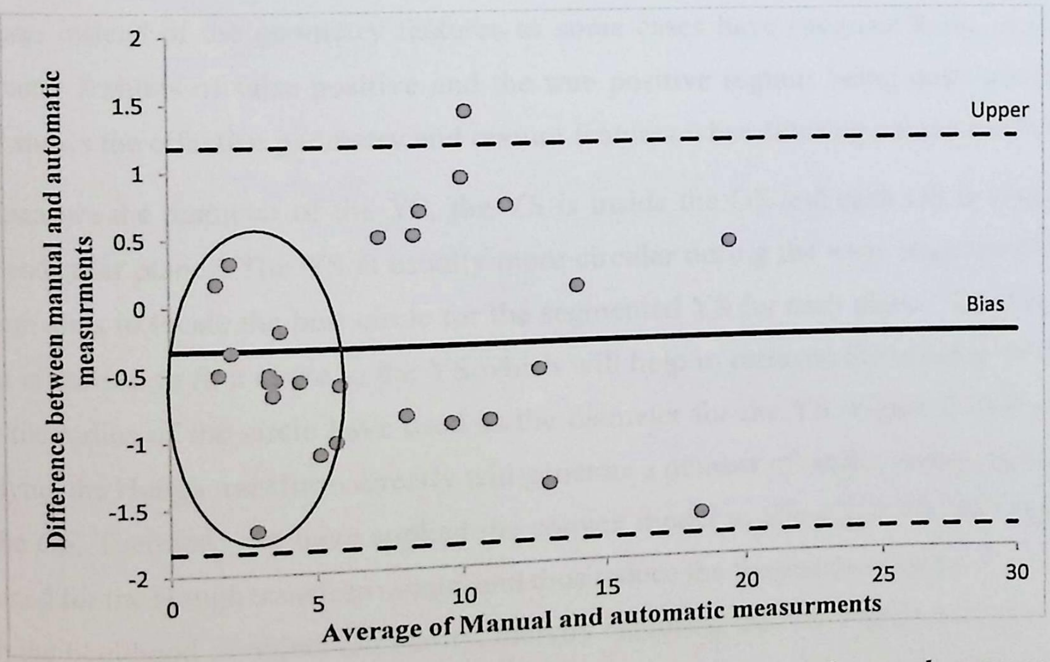



*Figure 5-25: Bland Altman correlation analysis of the automatic and manual measurements for batch 3 (GS+YS)*

*Table 5-5: Upper and Lower LOA for the Bland Altman analysis for Batch 1,Batch 2and Batch 3*

| | Batch 1 and Batch 2 | Batch 3 |
|---|---|---|
| **Upper LOA** | 4.951058931 | 1.23707 |
| **Lower LOA** | -2.414579318 | -1.86085 |

### 5.2.8 Discussion

Figure 5-26 illustrates the result of each step in the GS and YS segmentation of the pregnancy stage 2. The aim of the first level of the trainable segmentation is to capture the GS (see Figure 5-26-B and the second level of the trainable segmentation is to capture the YS (see Figure 5-26-E). We tried to segment Bothe the GS and YS using the one level trainable model, and as can be seen in Figure 5-26-D the model cannot capture the poor border of the YS. In addition, we have trained the model with three classes (Background, inside the GS, and the border of the YS); as can be seen in Figure 5-27, the three class confuse the model in terms of capturing the correct ROI due to the similarity between the texture of the YS border and some region in the background.

Figure 5-26-C shows the result of the filtering out the non-ROI stage. This stage was based on the trainable texture model to select the correct ROI. In this stage, we have used the texture features instead of the geometry features as some cases have irregular ROIs, as well as the geometry features of false positive and the true positive regions being quite similar. Figure 5-28 shows the effective geometry and texture features when filtering out the non-ROI.

To measure the diameter of the YS, the YS is inside the GS and each GS is viewed in two perpendicular planes. The YS is usually more circular during the early stages, and hence our system aims to locate the best circle for the segmented YS for each plane. The Hough Transform is utilised to fit a circle to the YS which will help to estimate the missing YS border, as will the radius of the circle have used as the diameter for the YS. Figure 5-26-F shows that applying the Hough transform directly will generate a number of circles pertaining to the shape of the GS. Therefore, we have applied the convex model to allow just the YS's border to be retained for the Hough transform stage, and thus reduce the frequency of objects in conjunction with the likelihood of object circularity, thereby capturing the YS. Figure 5-26 (G) - (I) show the effect of the convex on the result.

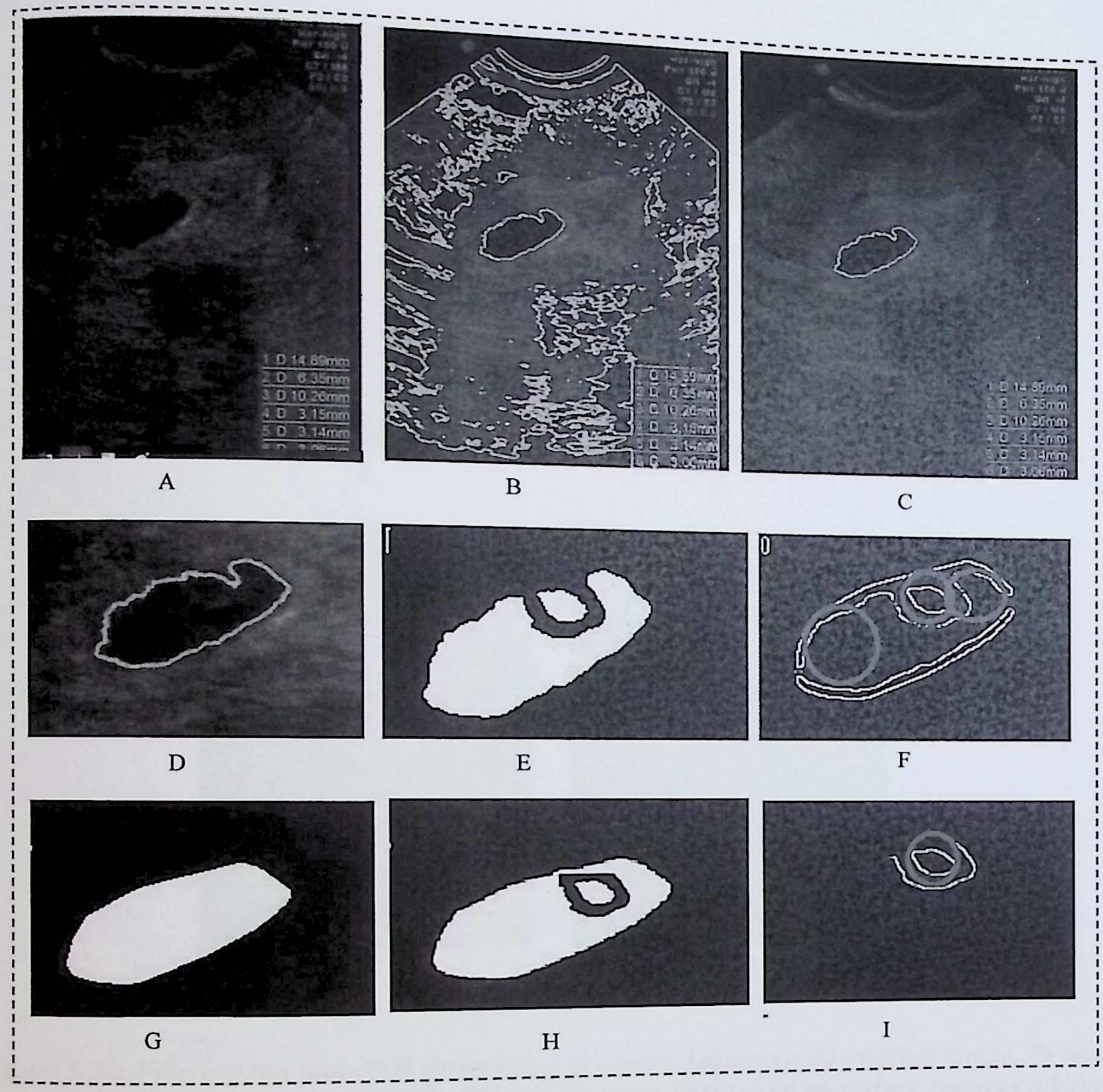


*Figure 5-26: GS & YS segmentation steps: A) original image (pregnancy stage 2), B)First Level Trainable Segmentation, C)filtering out non-sac, D) crop the ROI, E) second level trainable segmentation, F) circle Hough transform for the whole mask, G) the convex mask, H) extract the YS based on the convex mask and I) circle Hough transform.*

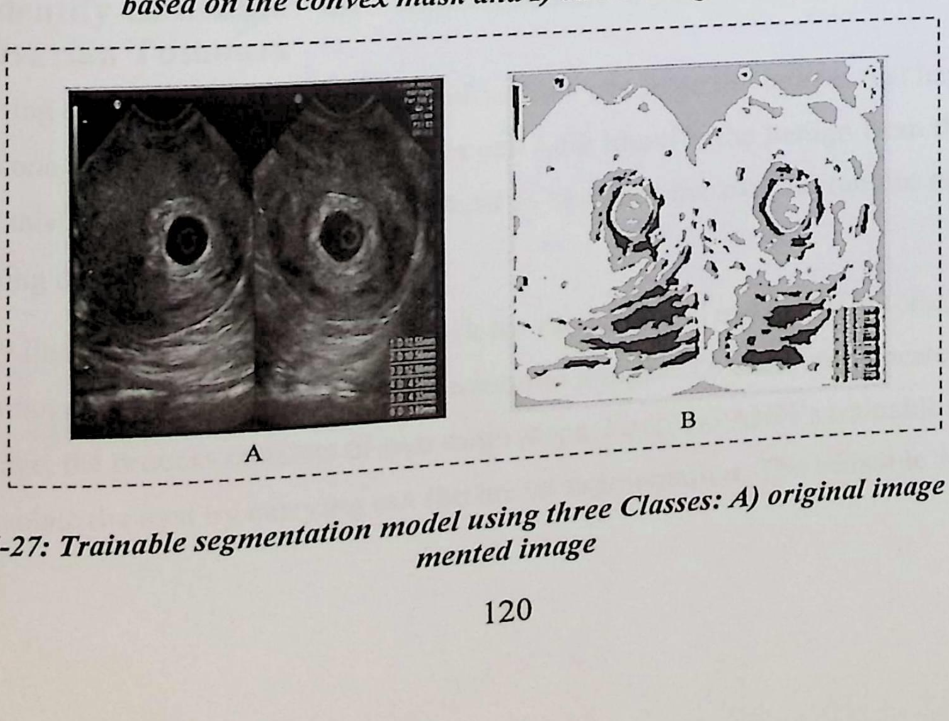


*Figure 5-27: Trainable segmentation model using three Classes: A) original image and B) segmented image*

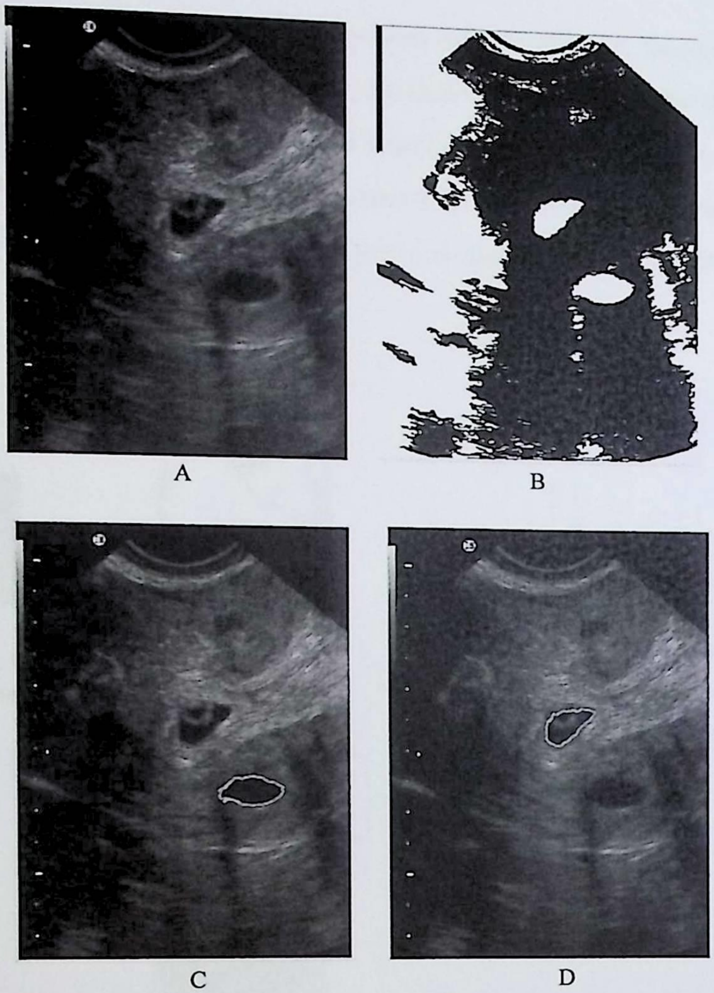



*Figure 5-28: Filtering out non-ROI (texture vs. geometry features): A) original image, B) binary image, C) filtering out non-ROI and D) filtering out non-ROI based on texture features (object detection).*

## 5.3 Proposal 2: Using Trainable Segmentation and Watershed Transform to Identify Unilocular and Multilocular Cysts from Ultrasound Images of Ovarian Tumours

Distinguishing ovarian images that have unilocular cysts from the images that have multilocular cysts is one of the important aspects that can help identify the benign from the malignant. Therefore, this section presents a new method to segment the ovarian tumour mass followed by calculating the number of cysts.

Figure 5-29 illustrates the process framework for the proposed method of automatic identification of ovarian cysts from static B-mode ultrasound images. To accurately locate a cyst within a given image, the process consists of two main steps. First, the ANN's trainable segmentation is used to isolate the cyst by carrying out the initial segmentation. The trainable ANN has used to generate a binary image by training the system on two classes. Second, the distance transform and watershed transform are employed to estimate the border. The distance transform can be used as this considers the object shape and interior. The minimum grey value for each object is labelled using the application of h-minima (marker-controlled) (Jung and Kim 2010) and the border estimation between the overlapped objects is carried out by watershed transform.

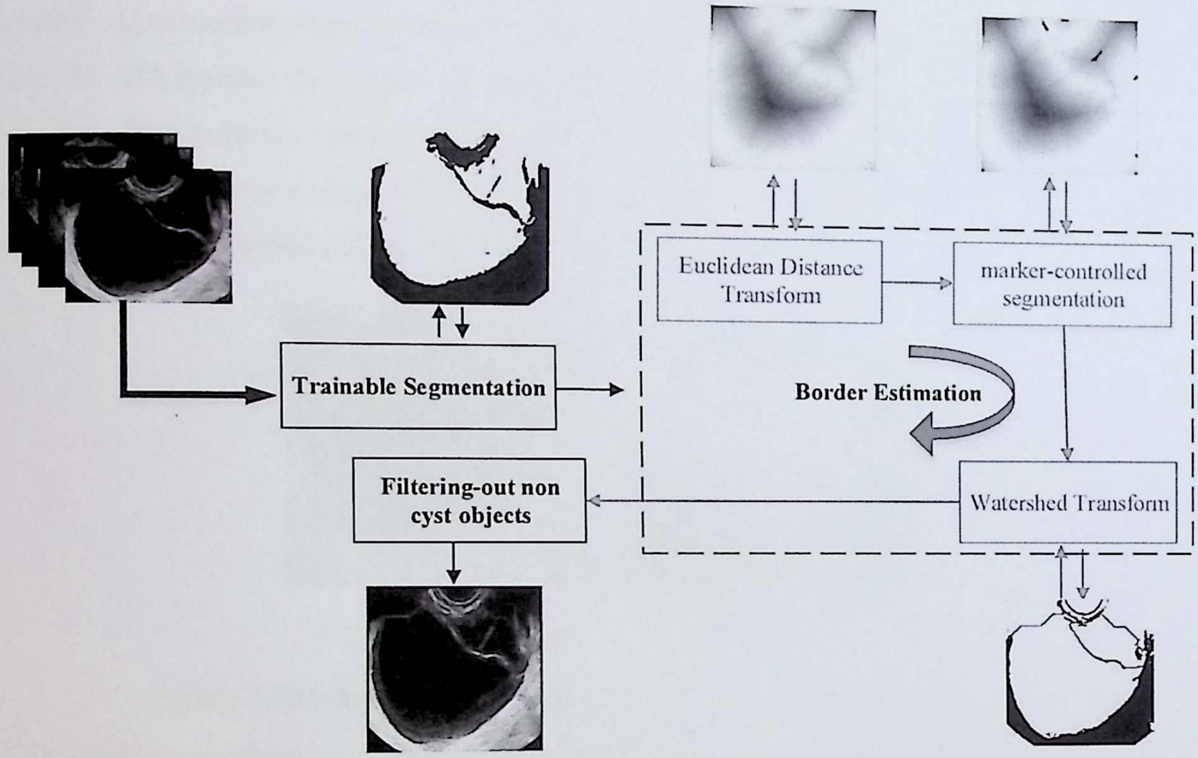


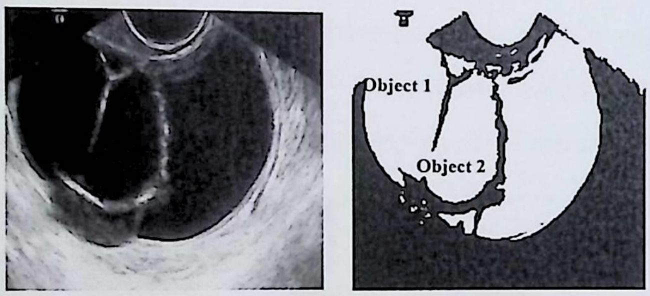

*Figure 5-29: Identifying unilocular and multilocular cysts from ultrasound Images based on trainable segmentation and watershed transform*

### 5.3.1 Trainable Segmentation

Trainable segmentation is intended to separate the region of interest (ROI), i.e., the tumorous area, from the rest of the image. This process starts by collecting a number of training images, each of which is then divided into a number of samples. A sample is a small square window of pixels of a particular size, e.g., 5x5, taken either from areas inside (Class 1) or from areas outside (Class 2) the ROI. A set of HOG features are then extracted from each window, forming a feature vector for the sample. A neural network classifier is then trained from this collection of labelled feature vectors. After completion of the neural network training, the image to be segmented is scanned pixel by pixel. A small window of the same size, e.g., 5x5, is taken around each central pixel whereby the HOG feature vector is extracted from the window and classified as being either inside or outside the ROI by the trained neural network. When a window is classified as being inside the ROI, the central pixel is set to one. Otherwise, it is set to zero.

### 5.3.2 Cyst Border Estimation

The estimation of the cyst's border is achieved by using the binary image produced by the trainable segmentation stage as input. It is possible that two objects may overlap (see Figure 5-30). To differentiate the shapes of the individual objects, all connected objects need to be separated. To accurately estimate the borders of the overlapping objects, this step of the proposed method involves a sequence of processing stages, starting from Euclidean distance transform, followed by H-minima, and finally the watershed transforms.




A) Original Image B) Binary Image

***Figure 5-30: shows the overlapping between object and the background***

- **Euclidean distance transform and H-minimum (Marker-control)**

To avoid overlap between the objects and estimate their borders, a subsequent process of segmentation is needed after initial image segmentation which is able to detect these aggregates. Popular methods for aggregate detection and segmentation include the distance transform method of aggregate associated binary images and watershed transform following the introduction of adequate markers. Binary image distance transform is determined in the following manner: every pixel x in set A, DT (A) is its distance from x to the complement of A:

$$DT(A)(X) = \min\{d(x,y), y \in A^c\} \tag{5-3}$$

A binary image's distance transform is, therefore, calculated by assuming that $A^c$ is the set of 1-valued pixels; this forms a greyscale image that can then be segmented using the watershed transform. However, unless the markers have been appropriately selected and applied, it has been shown that watersheds are prone to significant over-segmentation (Grau, et al. 2004).

Figure 5-31 shows the effect of a typical distance transform calculation. The first 'coarse' segmentation generates a binary image of two overlapping objects (Figure 5-31 (A)). However, the coarse segmentation does not separate overlapping objects, and as such it is possible to generate this binary image using the neural network's trainable. This binary image is then negated (Figure 5-31 (B)), before having the distance transform applied to generate the greyscale image in Figure 5-31 (C). It can be seen here that the black region in Figure 5-31 (B) that are the furthest from the white background appear as maxima. Despite this, several local maxima also appear as a result of figure morphology characteristics. By complementing the greyscale image in Figure 5-31 (C), Figure 5-31 (D) is generated giving a white background, and where the former maxima appear as minim. This method is described as an inner distance transform. The watershed transform is then applied to Figure 5-31 (D) to generate separation between the overlapping objects. It is possible to complement the region to be segmented outside of the aggregate to obtain adequate external markers. Over-segmentation causes the generation of the specious minima, and therefore it is necessary to ensure the construction of adequate markers.

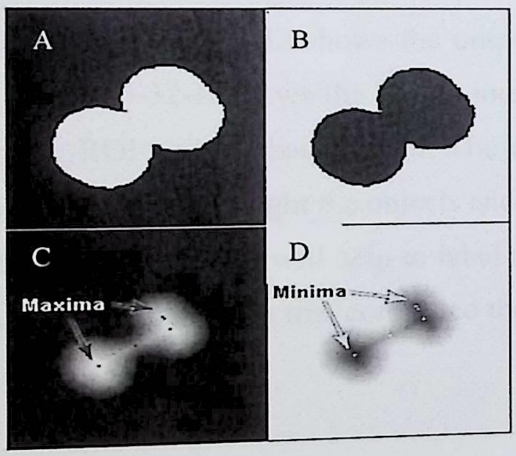



***Figure 5-31: A) Binary image from overlapping objects, B) complemented binary image, C) distance transform of the image in (B), and (D) complemented distance transform. Notice the maxima and minima indicated by arrows in (C) and (D).***

It is possible to use several greyscale morphological functions when setting the inner marker process; this will control the negative effects of specious minima prior to the application of the watershed transform. These morphological functions include:

- The imposition of minima at specific points by inserting a $-\infty$ value in a specific position to eliminate the existence of local minima at other points within the greyscale image. When these are placed in appropriate points, effective markers are created.
- From the result of the inner distance transform, the H-minima can be applied to eliminate all of the minima with a depth that is equal to or less than a given positive threshold.

This will reduce the remaining minima's depth within the magnitude of the threshold, therefore local minima can be eliminated by using an appropriate threshold and the generated regional minima can be used as effective markers.

The implementation of this method can be completed using a morphological sub-geodesic reconstruction $\nabla D$ of the surface intensity of the image $f$. $h$ determines the surface increased by the threshold, and the structuring element that defines the connectivity is $D$ (Soille 2013).

$$HMIN_{h.D}(f) = f(\nabla_D(f + h) \qquad (5\text{-}4)$$

Despite this simplistic description, the actual task of determining appropriate H-minima transform thresholds and identification of a suitable place to impose a minimum is complex and non-trivial. It is easy to misplace the markers, thereby allowing specious minima to remain. Additionally, regional minima can merge, thereby preventing marker isolation. Through the analysis, it can be seen that the actual task requires the elimination of specious minima as well as the minimum in the blob connections. This needs to be done whilst simultaneously maintaining the isolation of the two minima that reside at the approximate centre of the overlapping objects that need to be segmented. Figure 5-32 shows the output of the Euclidean distance transform and H-minimum. Figure 5-32-B shows the binary image and it may be noted that there is an overlap between the ROI and the background. The distance transform in Figure 5-32-C) is applied to the binary image to highlight the objects and makes the area that links the objects with the background very poor, which will help to label the real objects. H-minimum has used to label the objects and ignore the area that connected them with the background (see Figure 5-32-D).

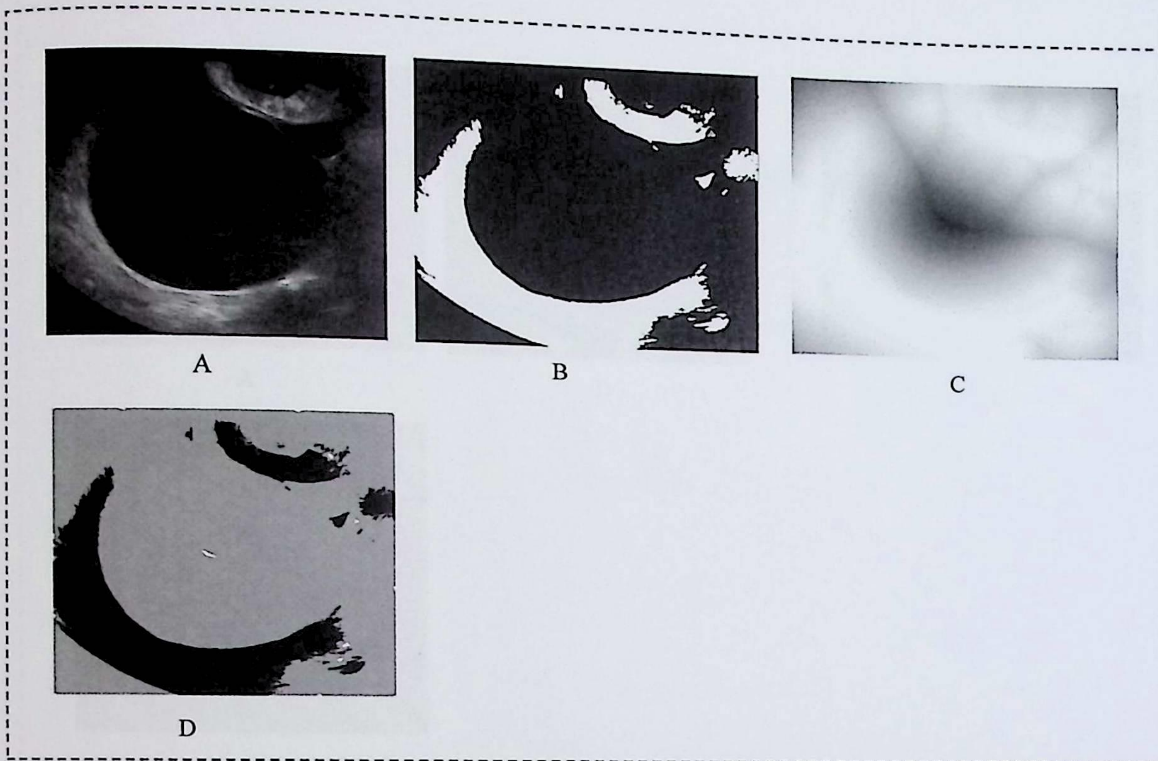



*Figure 5-32: An illustration of the Euclidean distance transform and H-minimum: A) original image, B) binary image, C) distance transform and D) label the objects (H-minimum)*

### 5.3.3 Watershed Transform

In order to achieve initial separation between the overlapped objects, this work applies the fast immersion-based watershed transform on the output of the gradient-weighted distance transform. As illustrated in Figure 5-33, the watershed transform helps to close the border of the cyst, where Figure 5-33-A shows the output of the previous stage, which is the H-minimum, to label the object. These labels are used as an initial seed for the watershed transform and, as based on these seed points, the watershed transform will grow up. Figure 5-33-B shows the watershed lines that represent the objects' border. In Figure 5-33-C we have removed the watershed lines that are not present on the border of the object based on the labelled image A.

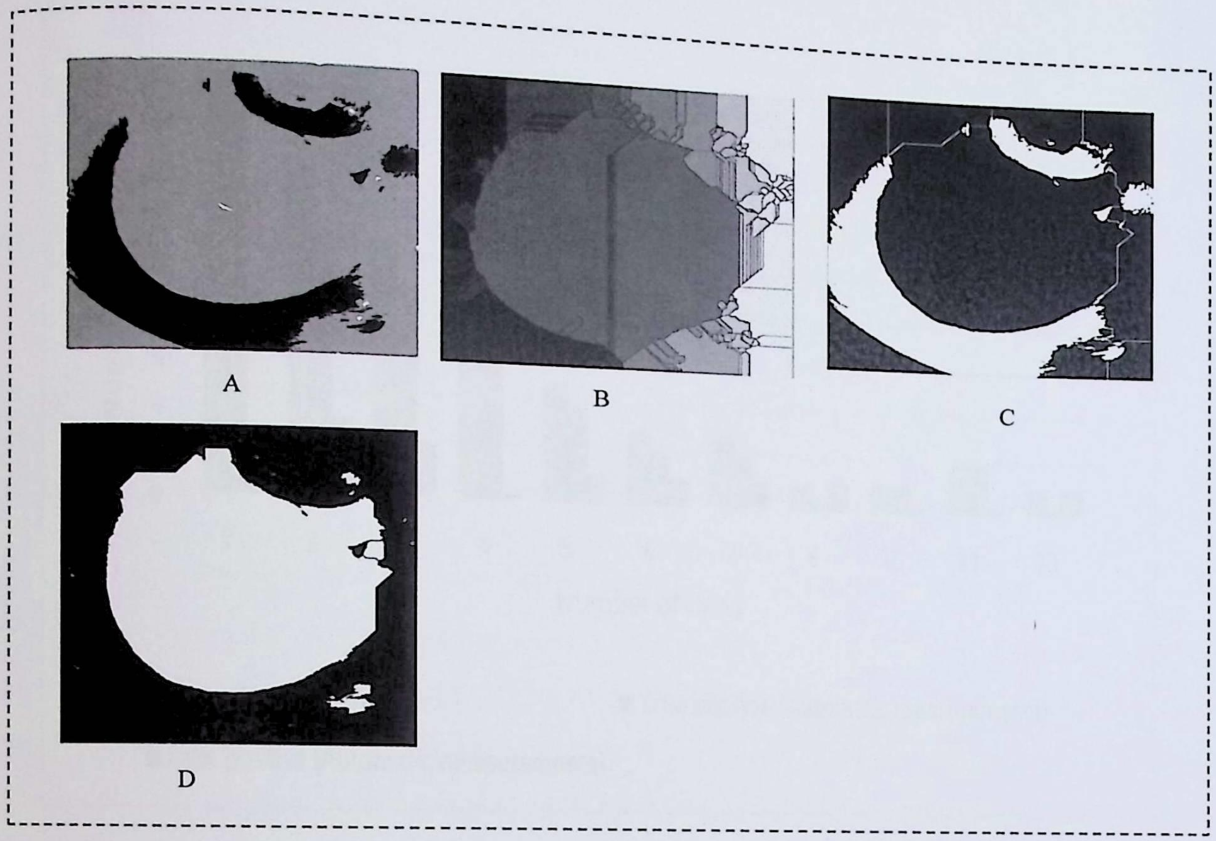



*Figure 5-33: An illustration of the watershed transform: A) shows the output of the previous stage ((H-minimum)), B) the watershed transform output, C) filtering out the lines to get the border only and D) output mask with clear ROI.*

### 5.3.4 Results and Discussion

#### A- Results

To obtain a ground truth to evaluate the effectiveness of our proposal, a domain expert was asked to manually count the number of cysts in 65 ultrasound images. This was then used as a ground truth to compare the automatic measurements against. We found that 54 out of 65 images were accurately segmented and produced the correct number of cysts, whilst 11 images produced an incorrect number of cysts. Therefore, the success rate is around 83%. Figure 5-34 shows a more detailed comparison between manual and automatic measurements in relation to the number of cysts in each ultrasound image. The blue bars of the automatic counting represents the true positive (i.e., the number of images that have a correct cyst count) whereas the red bars represents the false positive (i.e., the number of images that have inaccurate cyst count).

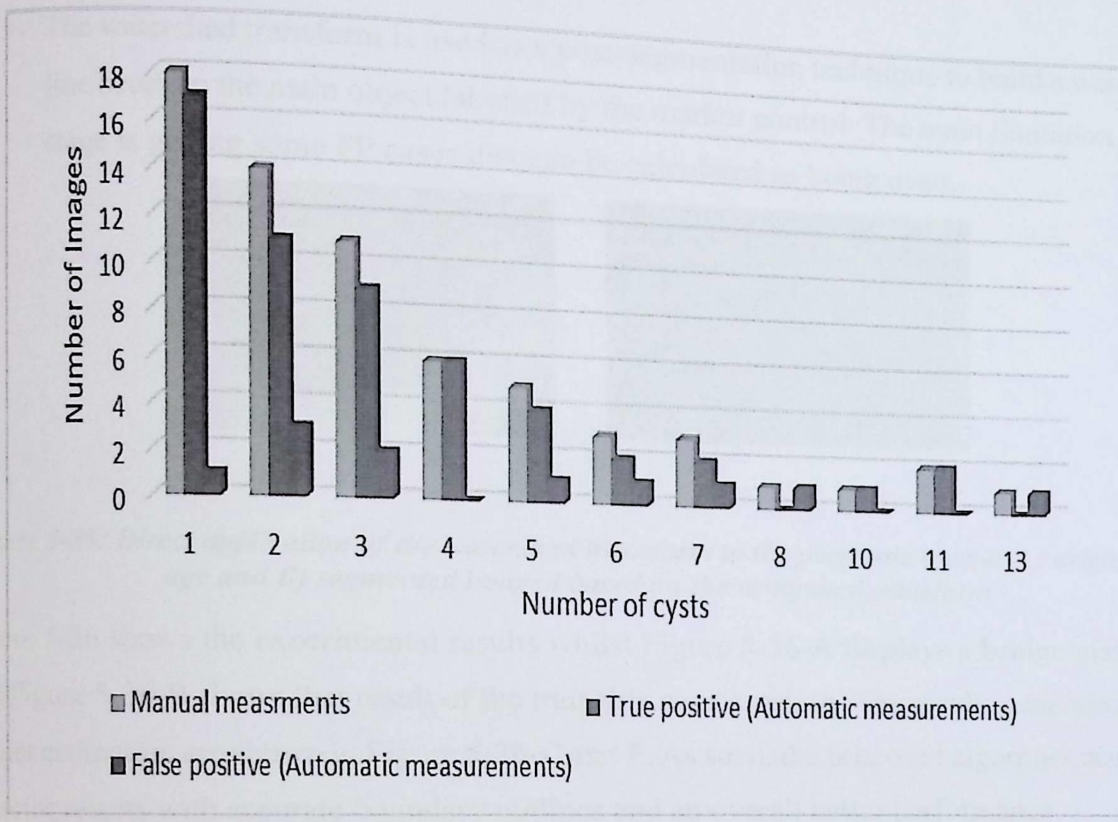



*Figure 5-34: Cyst counting - Automatic vs. Manual.*

**B- Discussion**

As mentioned previously, this work aims to combines trainable segmentation with the watershed transform to binarize the image and reconstruct the missing border. The watershed transform is a morphological segmentation method that is used to build the watershed line between objects. This algorithm is extremely sensitive to any noise (Meyer 1994). Therefore, applying the watershed transform directly to the ultrasound image would result in over-segmentation (see Figure 5-35). In addition, the limitation of using the trainable segmentation alone to segment the ultrasound ovarian tumours is the overlap between the cysts (i.e., two cysts have the same border) due to the poorly defined border between the cysts. To avoid these problems, we have a proposed solution that involves three steps:

1- Trainable segmentation method to binarize the image and catch the poor border. The output of this stage is a binary image. The main limitations of this stage are, first, in some cases we saw overlap between cysts (see Figure 5-36-B). Second, it requires a number of samples to train the model which should be taken from different cases.
2- The distance transform is applied to the binary image to highlight the objects in order to identify the overlapping area between the objects. The marker-control is used to label the main objects in the image.

3- The watershed transform is used as a post-segmentation technique to build a watershed line between the main object labelled by the marker control. The main limitation of this stage is getting some FP cases that can be calculated as being cysts.

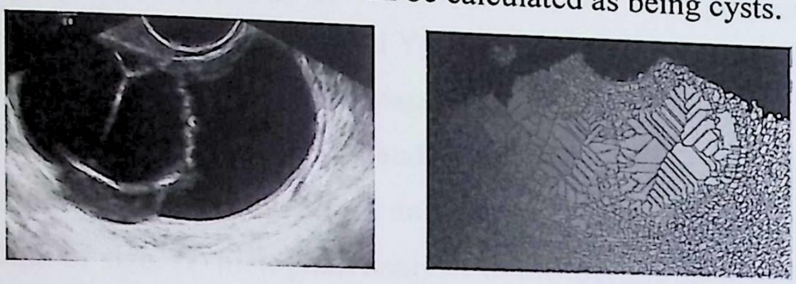



***Figure 5-35: Direct application of the watershed transform to the greyscale image: A ) original image and B) segmented imaged based on the watershed transform***

Figure 5-36 shows the experimental results whilst Figure 5-36-A displays a benign mass ROI and Figure 5-36-B shows that result of the trainable segmentation. The results generated using border estimation are shown in Figure 5-36-C and F. As such, the proposed algorithm displayed superior results with accurate boundary outlines and an overall better performance.

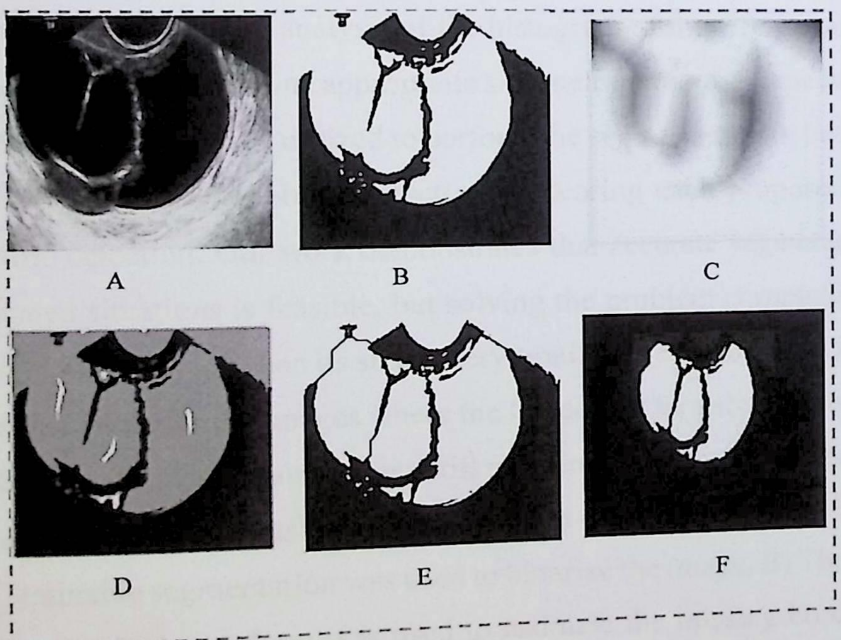



***Figure 5-36: Segmentation result, (A) original image, (B) trainable segmentation, (C) distance transform, (D) H-minima, watershed transform, (E) segmented image.***

## 5.4 Summary

For this investigation, the closely associated topics of segmentation of the GS and YS with respect to automatic identification of miscarriage from B-mode US scanning images have been examined. It was suggested that the quality of the image with regards to such matters as speckling, inhomogeneity of the ROI, the irregularity of the ROI, poor border, false positive objects as well as similarity between the texture of the ROI and the background all pose technical challenges in relation to correct ROI segmentation.

1. An innovative multi-level segmentation approach has been proposed to achieve two objectives: **1**- Segmenting the GS and YS, followed by measuring the MSD to identify instances of miscarriage in the early stages of pregnancy. To achieve this, the proposed approach involves: A) segmentation and measurement of the GS, B) filtering out non-sac objects using the texture features and SVM model, (C) accurate sac segmentation for estimation of size, D) histogram analysis to identify the stage of pregnancy, E) segmentation of the YS, and F) locating the YS and measuring its size. The initial segmentation process (A) involves using a trainable segmentation procedure founded on the histogram of oriented gradients to segment the GS. Then, (B) is about reducing the false positives and selecting the correct object, for which we have proposed a method based on the SVM to reduce the number of unwanted objects and detect the ROI. (C) is utilized to mitigate the under-segmentation problem, whereby the region growing method has been applied, followed by the measuring the MSD. Finally, segmentation (D) identifies the pregnancy stage by analysis of the histogram of the GS. (E) and (F) are used to identify the YS and determine appropriate size measurements, respectively. A trained neural network classifier was utilized to perform the segmentation in both A and C. The SVM classifier is used for object detection by scoring each proposed region with a class-specific detection. Our work demonstrates that accurate segmentation of the GS in complicated situations is feasible, but solving the problem completely is extremely difficult, if at all possible, when its size is very small. Our approach has a validity check embedded that excludes the images where the GS cannot be successfully segmented. In practice, this means that certain images still need intervention by a gynaecologist.
2. Identification of ovarian cysts from static B-mode ultrasound images. This process includes: A) trainable segmentation was used to binarize the image. B) The distance transform with watershed transform was used to estimate the missing border followed by calculating the number of the cysts in the images.

Based on our investigations, machine can help to avoid several limitations i.e., reduce false positives, catch poor borders, filtering out objects that are not of interest as well as the training samples and the classifier being used to help to reduce the problems we faced with the more traditional segmentation in chapter 3 (over- and under-segmentation).

# Chapter 6. Improvements on Trainable Segmentation

In the previous chapters, we introduced a multi-level trainable segmentation model to address some challenges related to the ultrasound image segmentation. This chapter is concerned with improving the trainable segmentation to further address two key limitations facing the proposals in chapters 4 and 5. Limitation 1 is the under-segmentation problem of the traditional Region Growing (RG) technique used post-segmentation in section 5.2.3. The ROI of the ultrasound image has a different texture, and it can be argued that finding a suitable condition (threshold) to stop the growth of the RG based on pixel intensity is not sufficient to capture the correct edge, and, therefore, we cannot get the whole ROI as explained in Section 6.1.1. Limitation 2, on the other hand, is related to the similarity between the texture of the background and that of the ROI, which will result in an over-segmentation problem where the ROI's border is not clear. Therefore, classifying the pixels or fixed size window based on their local information is not sufficient, as discussed in Section 6.1.2. In this chapter, we address the above limitations by proposing two possible solutions.

1- Proposal 1: Enhancing the traditional RG using texture features and ANN (Trainable Region Growing) to address limitation 1.
2- Proposal 2: Hybrid trainable model to identify difficult cases by combining the Cascade model with the trainable segmentation method proposed in chapter 3 to estimate and crop the ROI (object detection) of those difficult cases to address limitation 2.

The rest of this chapter is organised as follows. In Section 6.1 we have described the challenges that we discovered with, first, the region growing method that we used post-segmentation and, second, the proposed trainable segmentation models based on fixed window size. Section 6.2 presents the proposed solution in terms of enhancing the region growing based on the neural network classifier and texture features, and justifies why this enhancement is necessary. Section 6.3 is devoted to our work to improve the trainable segmentation to avoid the over-segmentation problem. Section 6.4 summarises the work carried out in this chapter.

## 6.1 Problem Statement

The main challenges facing ultrasound image segmentation are: 1) inhomogeneity of the ROI, which means the ROI has different textures, causing a problem known as "under-segmentation", i.e., parts of the ROI are not detected, and 2) in some cases, the ROI and the background have a similar texture which makes automatically finding the ROI very difficult (over-segmentation). The following subsections describe the two limitations in detail.

### 6.1.1 Limitation 1: Under-Segmentation - Inhomogeneous Region of Interest

Our objective in this chapter is to improve the region growing and make it suitable for the ultrasound images. In this section, we will explain the RG limitations that we faced in our previous proposed algorithms.

The key factors influencing the output of the RG are, first, automatic generation of the seed points or mask (initial segmentation). Second the selection of similarity criteria, where, the intensity values (either the grayscale scale levels or other measurement that can easily be calculated from the intensity, i.e. texture features) have used as a similarity criterion in the traditional RG. Third, the stopping rule definition for the RG, i.e., RG should stop growing when there is no further region that satisfy the similarity criteria and homogeneity to be involved in that region.

The initial segmentation (seed mask) has been addressed using trainable segmentation (see chapter 5). Therefore, this section will focus on the selection of the similarity criteria and the stopping rule that can help to control the growth of the RG.

Figure 6-1 shows the effects of the traditional RG using different ultrasound images. However, one can notice the under-segmentation problem, i.e., the whole ROI cannot be captured, as illustrated in Figure 6-1, where the first column represents the original images for different early stage pregnancy cases and the second column shows the under-segmentation problem after applying the traditional RG. Based on our investigation, has been found that the main reasons behind the problem are 1) different texture in the same ROI due to variation in the GS tissues or the speckle noise that comes from the prop movements, 2) different textures across different images and, therefore, determining a suitable threshold to control the RG is not easy. This means that segmenting the ROI based on the intensity as a similarity criterion is not sufficient when we have an inhomogeneous object of interest. To avoid this problem, a trainable region growing model has been proposed in section 6.2 that can help to control the growth of the region growing.

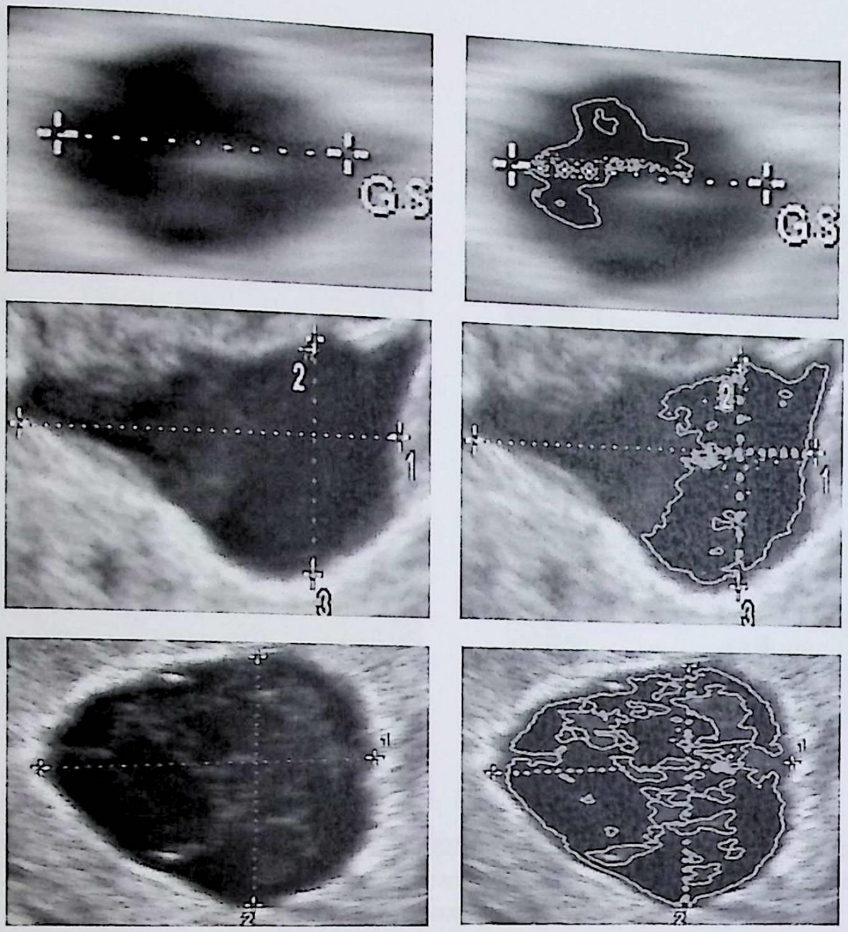



*Figure 6-1: Inhomogeneous ROI and the effect of the traditional RG*

### 6.1.2 Limitation 2: Over segmentation – no clear object border

Automatic ovarian tumour segmentation is still an open and challenging problem (Sohail, et al. 2010) (Hamid 2011). In fact, a good number of US images in our dataset have made the automatic segmentation of the ovarian tumour a particularly challenging task that might hinder the feature extraction. Two limitations related to the border (poor and missing border) have been covered in chapter 5 by proposing a new trainable segmentation model followed by the watershed transform as a post-segmentation stage to segment and calculate the number of cysts. This algorithm works with images where there is a difference between the texture of the ROI and the background. Figure 6-2 shows the effect of the trainable segmentation that we proposed in chapter 5. These kinds of image show a difference between the texture of the ROI and the background. Therefore, a trainable model based on a small fixed-size window might well work effectively.

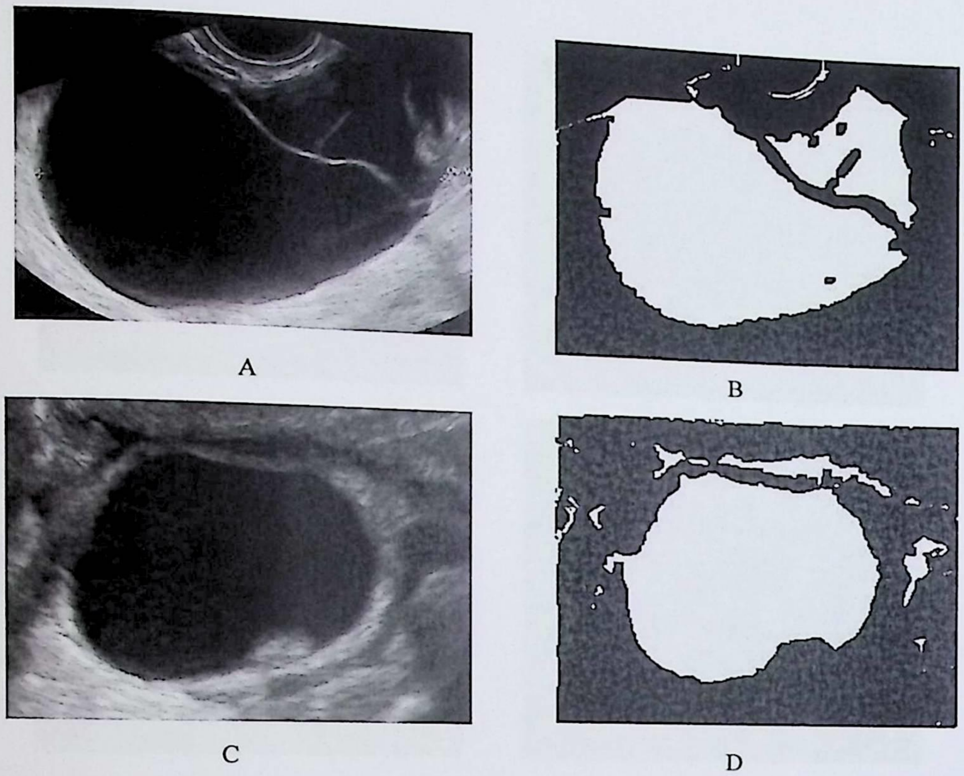


*Figure 6-2: Trainable segmentation based on fixed block size: A and C are the original images for different ovarian tumour cases, and B, D represent the segmented images using a trainable model based on a fixed small-size window.*

In Figure 6-3 we have applied the same proposed algorithm on more complex images and ended up with an over-segmentation problem due to 1) the overlap between the texture of the ROI and the background, 2) there is no clear border that can help to extract the ROI, 3) the proposed trainable segmentation model scans the image with a small fixed-size window that are not suitable for detecting the edge of the ROI. Figure 6-4 shows the similarity between the ROI and the background. In this figure, we have taken a fixed-size window from both the ROI and the background. The histogram for each window shows that there is only a small change between the two histograms and so it is difficult to build a classifier that can distinguish the window that belongs to the ROI from any others. Therefore, we have proposed an adaptive hybrid model by combining the model proposed in chapter 5 with the Cascade model to deal with both cases (cases shown in Figure 6-2 and Figure 6-3) and for which more details are given in section 6-3.

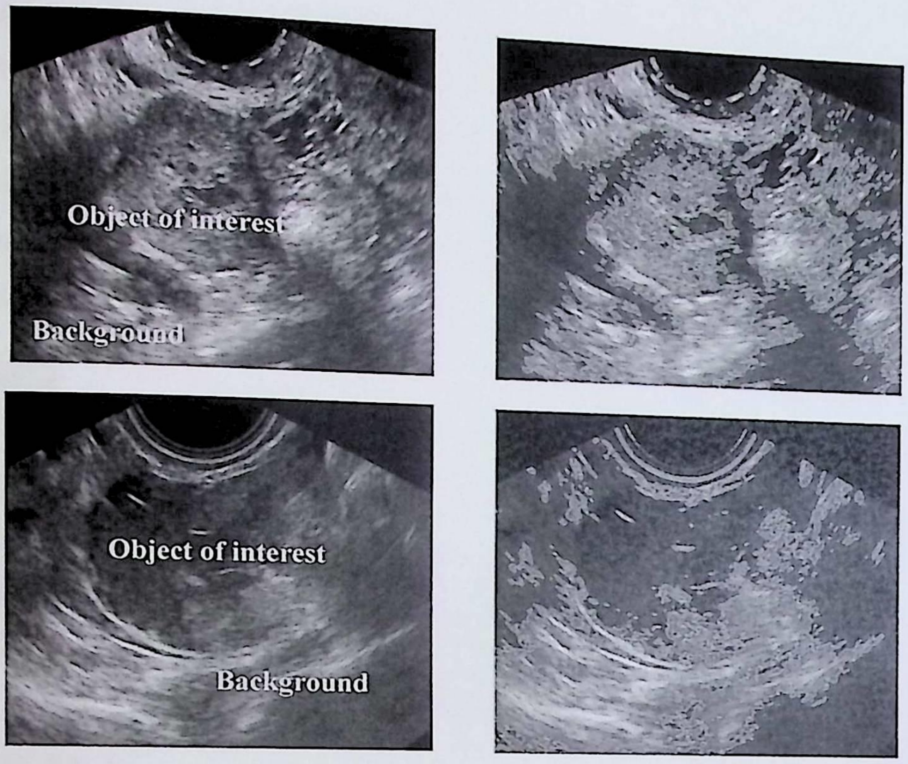



*Figure 6-3: Ovarian tumour samples show the similarity between the ROI and background and the over-segmentation problem.*

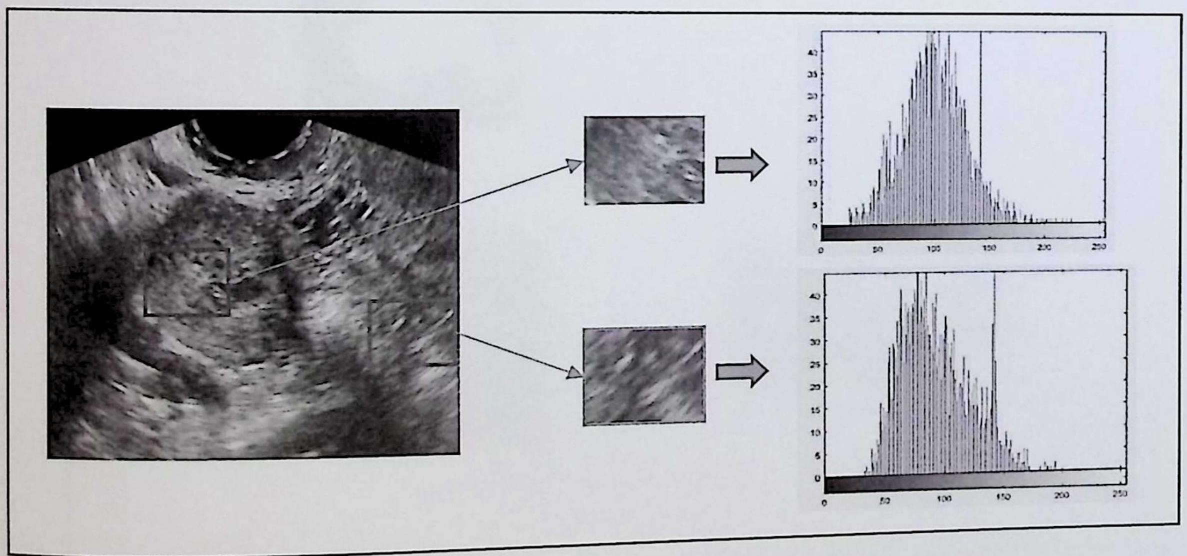


*Figure 6-4: The similarity between the texture of the ROI and the background.*

## 6.2 Proposal 1: Trainable Region Growing Model based on Neural Network

Figure 6-5 shows the proposed trainable region growing model stages. We argue that the proposed model addresses the limitations explained earlier in section 6.1.1. The similarity criterion used in this model is that of texture features instead of the greyscale intensity. Using the greyscale intensity (traditional RG) works well when we have a smooth ROI, i.e., this region has the same level of the intensity as can be seen in Figure 6-6 (A). In this case, we can easily set up a threshold that captures the difference between the seed point and the pixels around it. By contrast, the ROI in Figure 6-6 (B) has a texture inside the ROI similar to the background, and, therefore, finding a threshold to segment the whole ROI can be challenging because using a high threshold will result in the over-segmentation problem, whilst a low threshold leads to an under-segmentation problem. Therefore, using the texture features based on block instead of colour intensity can help to avoid this problem.

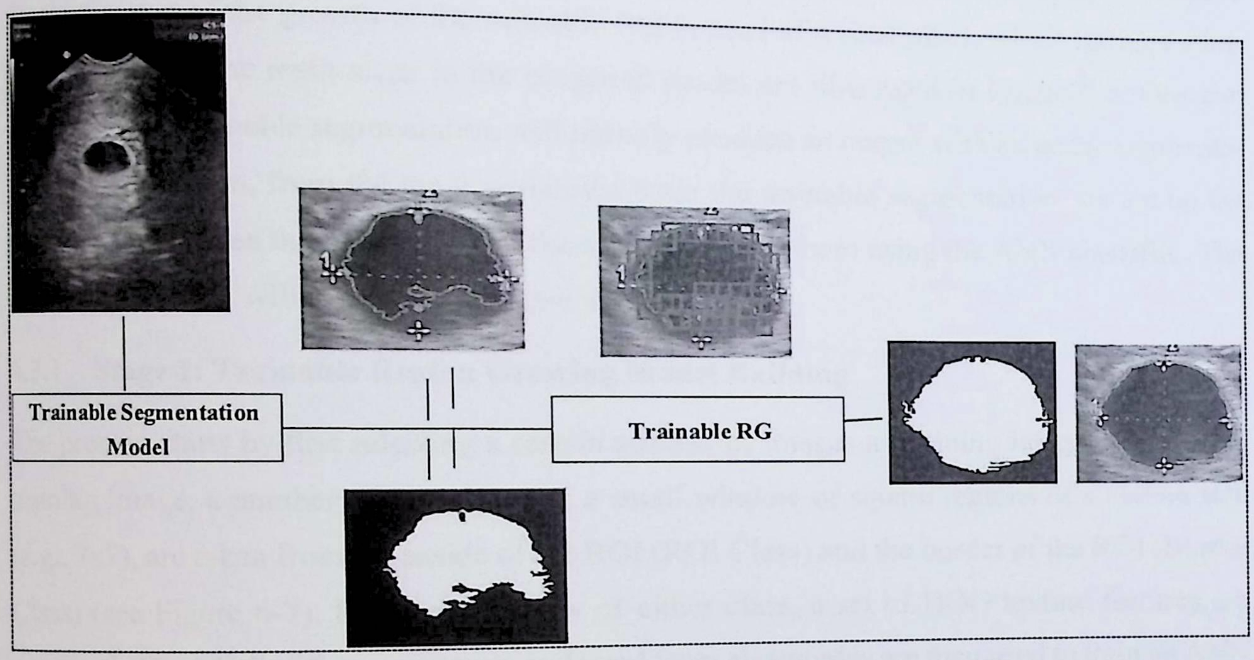



*Figure 6-5: Trainable region growing model*

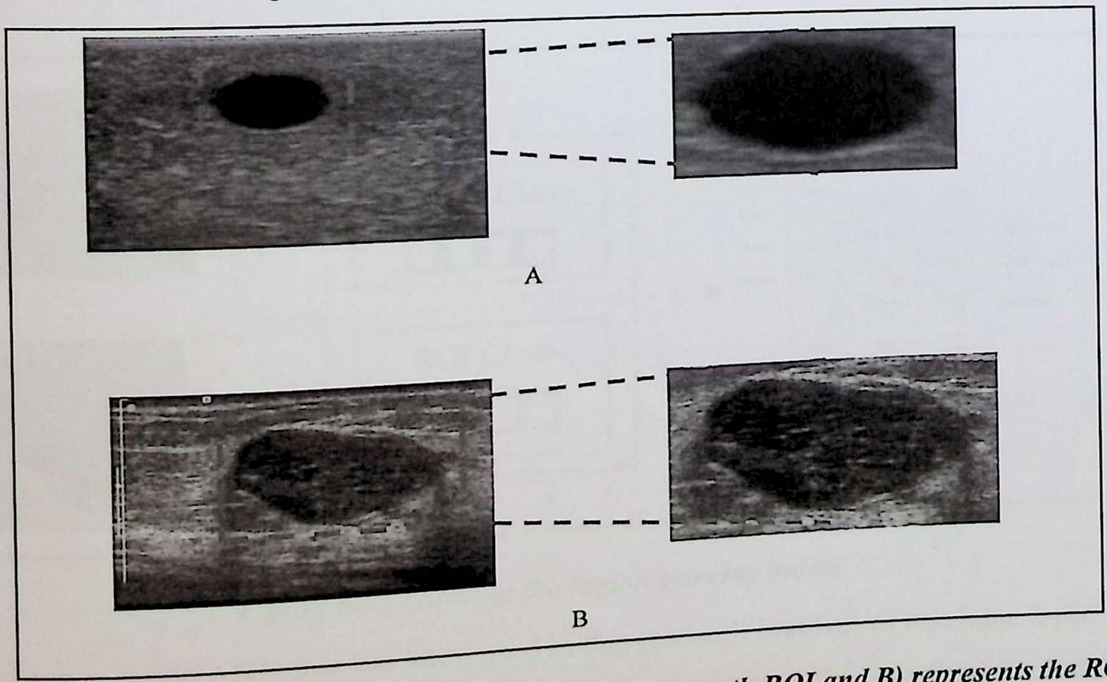



*Figure 6-6: Different cases of the ROI: A) represents the smooth ROI and B) represents the ROI with variance intensity.*

On the other hand, the stopping rule (threshold) is one of the limitations facing the traditional RG. Therefore, in the proposed trainable RG, the ANN classifier is used as a stopping rule. The proposed model includes two main stages. The first stage is implementing the model based on a number of samples to train the ANN classifier. The second stage is to classify the blocks using the trained RG model. This stage uses the mask that has been generated using the trainable model (see chapter 5) as an initial segment (seed mask). The generated mask will help to locate the start of the growth of the trainable RG instead of a seed pixel, which reduces computational cost. The main steps in the proposed model are illustrated in Figure 6-5. One can notice that the trainable segmentation will initially produce an output with an under-segmentation problem. Then, from the mask generated from the trainable segmentation, we set up the trainable RG to scan the blocks around the mask and label them using the ANN classifier. The following sections will describe the stages in detail.

### 6.2.1 Stage 1: Trainable Region Growing Model Building

The process starts by first selecting a certain number of images as training images. From each training image, a number of samples, i.e., a small window of square regions of a certain size (e.g., 7x7), are taken from the inside of the ROI (ROI Class) and the border of the ROI (Border Class) (see Figure 6-7). For each window of either class, a set of HOG texture features are extracted. The labelled feature vectors collected from all samples are then used to train an ANN classifier.

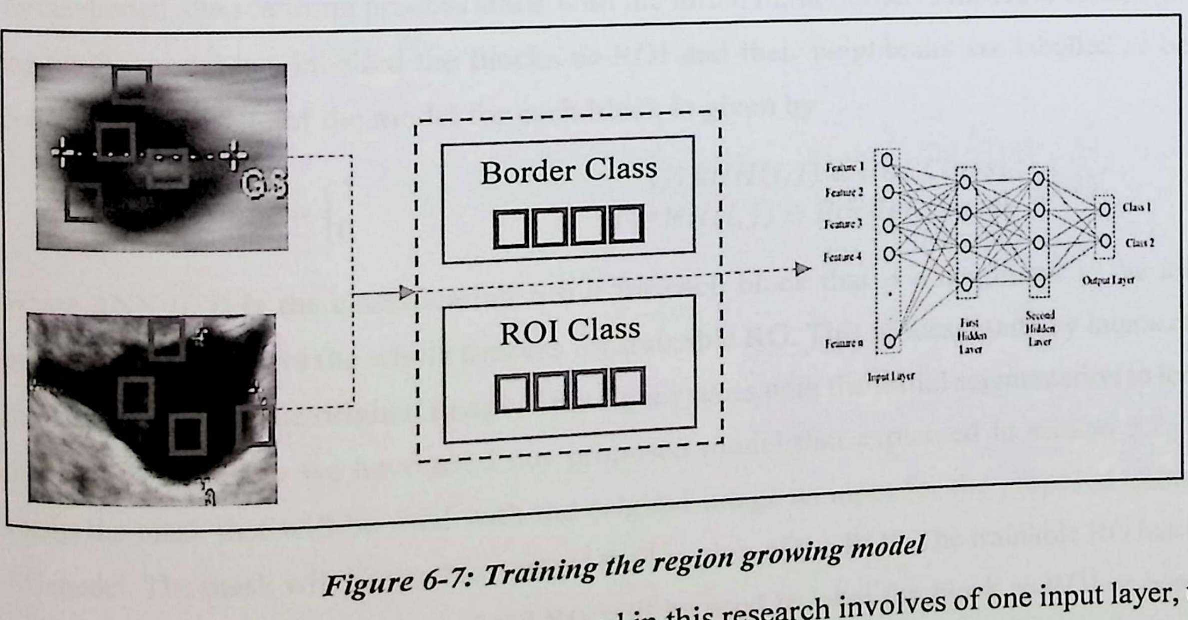



*Figure 6-7: Training the region growing model*

The back-propagation ANN that has been used in this research involves of one input layer, two hidden layers and one output layer. That classifier used to control the growth of the trainable RG. The features and the labels of the training samples are feed into the ANN to train it. The aim of the training stage is to find a mapping between the HOG vector features and the sample labels. This mapping is accomplished by adjusting the value of the weights $w_j$ using the learning algorithm. In this study, the input features are a vector of HOG values and the target vector, "whose dimension is equal to the number of target tissues in the image is binary", present the class label to which the block should be labelled. As described above, we should adjust the weight $w_j$ first to minimize the error which presents the difference between target and the observed vector. After adjusting the weights for the neural network, the region growing model is ready to work as a segmentation method to label the windows based on the local texture information.

### 6.2.2 Stage 2: Test the Trainable Region Growing Model

Once the trainable region growing model is built, each image from the testing set is used as an input with the seed mask for the trainable region growing model. The seed mask will be used to detect from where the proposed model should start. After selecting the seed mask automatically, the algorithm scans the image block by block (with an overlapping range) around the seed mask. For each block, the texture feature of the region is extracted and classified by the trained region growing model as being inside the ROI, or the border of the ROI. If the block is inside the ROI, the central pixel is labelled as inside the ROI; otherwise, it is labelled as the border of the ROI.

As mentioned, the scanning process starts with the initial mask border. This region keeps growing till the model has labelled the blocks as ROI and their neighbours are labelled as being border. The predicate of the model for each block is given by

$$predicate = \begin{cases} 1 & if|ANN(i,j) \in ROI\ Class| \\ 0 & if|ANN(i,j) \in Border\ Class| \end{cases} \quad (6\text{-}1)$$

Where ANN $(i, j)$ is the classification result for each block that is a neighbour to the initial mask. Figure 6-8 shows the whole process for trainable RG. This process starts by intersecting the initial mask with the original image. That figure starts with the initial segmentation to locate the ROI, in this stage we have used our proposed model that explained in section 5.2. The entails the mask that will be used with the original image as input for the proposed trainable RG model. The mask will be used to control the location of the ROI. The trainable RG has two stages, training and testing. The trained RG will be used to label the block as ROI or border. Figure 6-9 shows the entire process by applying the trainable RG on a GS ultrasound image.

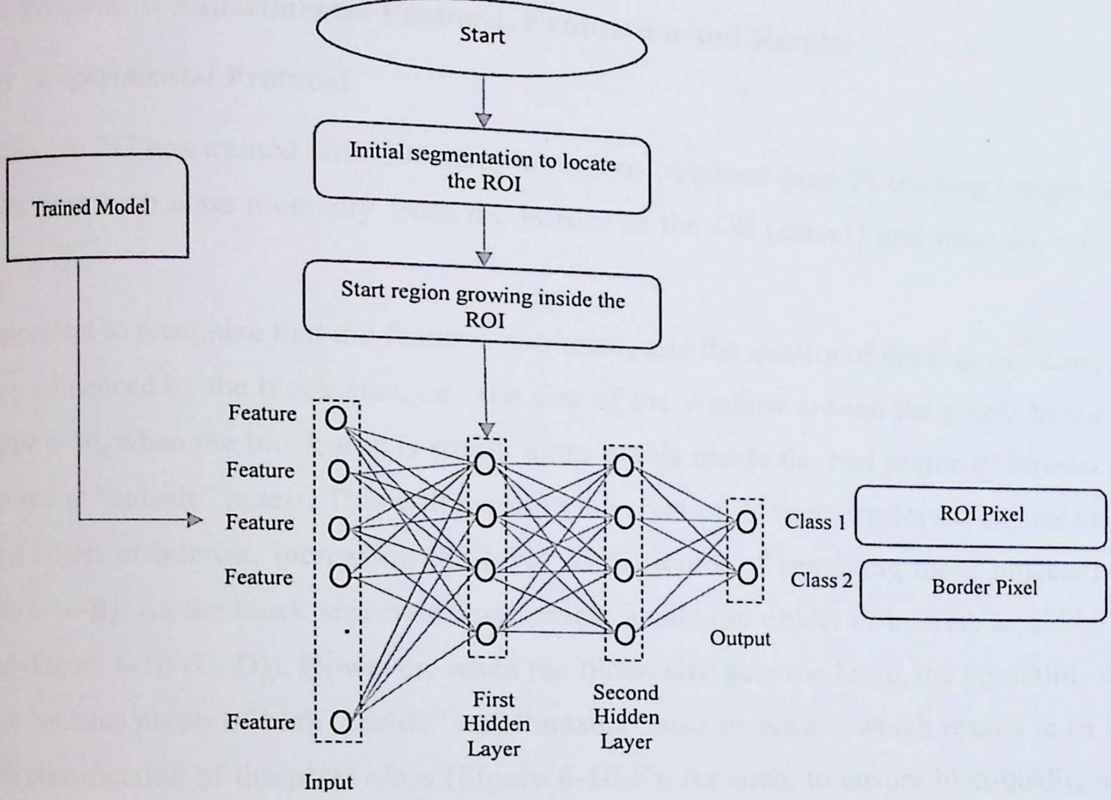



*Figure 6-8: Trainable region growing process*

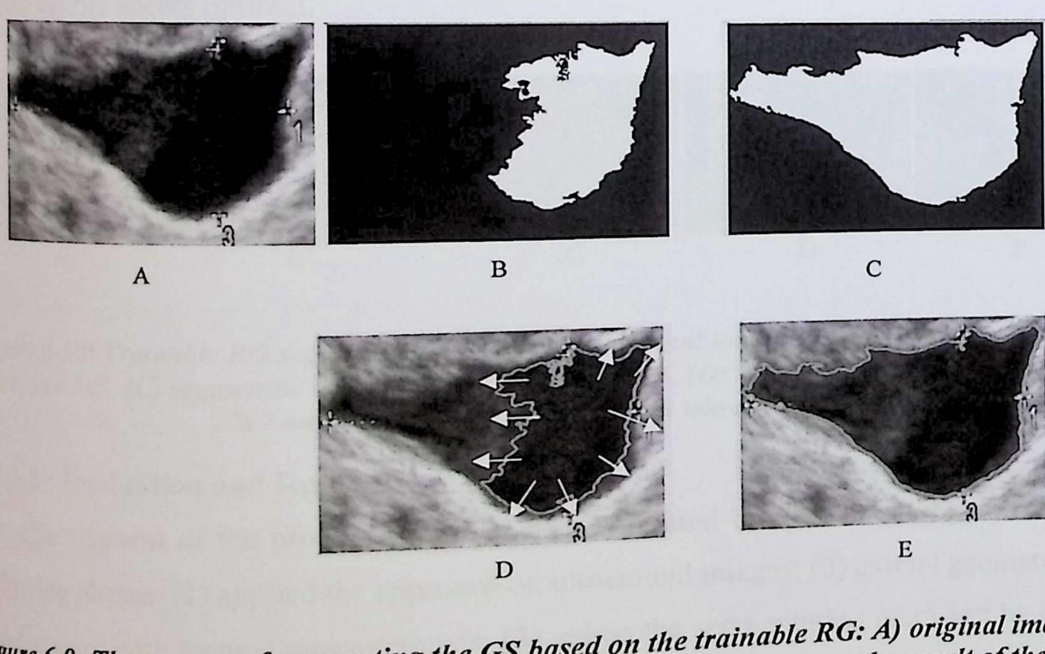



***Figure 6-9: The process of segmenting the GS based on the trainable RG: A) original image, B and C) represent the initial segmentation (initial mask); D and E) represent the result of the training RG.***

### 6.2.3 Proposal 1: Experimental Protocol, Evaluation and Results

#### 6.2.3.1 Experimental Protocol

The trainable RG was trained with 500 sample regions obtained from 25 training images. 250 samples from each class manually from the border of the GS (class1) and from the regions inside the GS.

It is important to recognize that the features extracted, and the quality of the segmentation, are directly influenced by the block size, i.e., the size of the window around the pixel. As shown in Figure 6-10, when the block size is small, more pixels inside the real region of interest are considered as "outside" pixels. This will result in the creation of more irrelevant objects inside the real object of interest, increasing the level of difficulty of removing those objects later (Figure 6-10-B). As the block size increases, pixels inside the object of interest would be removed (Figure 6-10-(C, D)). However, when the block size gets too large, the possibility that a block contains pixels of both "inside" and "outside" also increases, which results in an imprecise classification of the pixel class (Figure 6-10-E). As such, to ensure high-quality segmentation, the block size must be determined carefully. Based on our observation, a block size of 5x5 or 7x7 seems optimal.

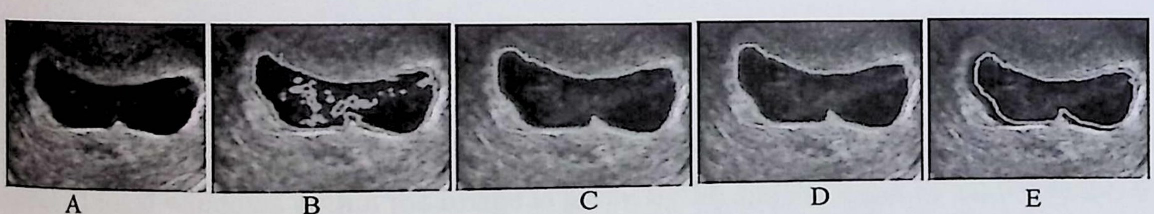



*Figure 6-10: Trainable RG segmentation result: (A) original image, (B) segmented image with window size 3x3, (C) segmented image with window size 5x5, (D) segmented image with window size 7x7 and (E) segmented image with window size 9x9.*

#### 6.2.3.2 Evaluation and Results

The effectiveness of the proposed method was evaluated by appropriate examination at the following phases: (1) applied the approach on ultrasound images; (2) extract geometry features and compare with manual measurements; (3) assess the segmentation precision by comparing the automatic and manual measurement results with the Structural Similarity Index, the Probabilistic Rand Index, the Variation of Information metric and the Global Consistency Error (Martin, et al. 2001) (Al-Fahdawi, et al. 2016).

**The Structural Similarity Index (SSI):** Measurement of the structural similarity index (SSI) between a segmented image (X) and a ground-truth image (Y) can be performed based on the segmentation quality assessment algorithm underpinning the SSI. Luminance, contrast and structure between X and Y in a local window are the elements subjected to measurement. The equation for the measurement is:

$$SSIM(x,y)\frac{\left(2\mu_x\mu_y+C_1\right)}{(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2+C_2)} \qquad (6\text{-}2)$$

In the above, the mean intensity and standard deviation of $x$ are denoted by $\mu_x$, while the mean intensity and standard deviation of $y$ are denoted by $\mu_y$; the covariance measure for x and y is given by $\sigma_{xy}$, while stability is kept by the small constants $C_1 = (k_1L)^2$ and $C_2 = (k_2L)^2$ when the value of $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ is close to zero. Furthermore, the dynamic range of the pixel values (255 for 8-bit grayscale images) is denoted by L, while $k_1$ and $k_2$ have values of less than 1. The $k_1$ and $k_2$ supplied in this study are set at the default value of 0.04, while L was set at 100. The movement of each pixel across the entire image and in each step of the local measurements is aided by the fact that the local measurements are within the local (8×8) square window. The calculation of the SSIM is undertaken in the local window and the equation employed for determining the mean of SSIM (MSSIM) to obtain the general quality measure of the whole image is:

$$MSSIM(X,Y) = \frac{1}{M}\sum_{i=1}^{M} SSIM(x_i, y_j) \qquad (6\text{-}3)$$

In the above, the image contents at the i-*th* local window are denoted by $x_i$ and $y_i$, while the number of local windows within the image is given by $M$, and the range of the MSSIM is 0-1, where the higher the value the closer the similarity. The SSIM index map is determined through provision of a measurement in the local image quality over space, where the brighter the SSIM index map the higher the quality of segmentation.

**The Probabilistic Rand Index (PRI):** Determination of the number of fractions of distinct pixel pairs involves the use of the Probabilistic Rand Index (PRI) and is achieved between the segmented and ground-truth images with harmonious labels by creating an average of the ground-truth images to calculate the scale variation on the basis of human perceptions. The range of PRI values is 0-1, where the higher the value the closer the similarity.

**The Variation of Information (VOI) metric:** Measurement of automatic distances and manual segmentation distances on the basis of data available in every distance is commonly undertaken with the non-negative metric called Variation of Information (VOI). Entropy and reciprocal information enable this metric to achieve determination of the distance between two clusters of information. The VOI between the segmented image (S) and the ground-truth image (S') is characterised via equation 20. The lower the value of VOI the closer the similarity is. The formula for the calculation is:

$$VOI(S,S') = H(S) + H(S') - 2I(S,S') \quad (6\text{-}4)$$

In the above, VOI is in the range 0-∞, H denotes the entropy and I denotes the reciprocal information. In this study, (S) and (S' ) give the reciprocal information and its determination is achieved with the formula:

$$I(S,S') = \sum_{k=1}^{K} \sum_{k=1}^{K'} P(k,k') log \frac{P(k,k')}{P(k)P(k')} \quad (6\text{-}5)$$

In the above, the joint probability distribution function characterised by (*S*) and (*S'*) is denoted by *P*(*k*, *k'*), the marginal probability distribution function of (*S*) is given by *P*(*k*), while the marginal probability distribution function of (*S'*) is given by *P*(*K*).

### The Global Consistency Error (GCE)

The degree to which visualisation of the segmented image as fine-tuning of the ground-truth image is possible is indicated by the global consistency error. The segment taking the form of a series of pixels indicates the consistency of the segmentations, while the validity of segment (*S*) as a subset of segment (*S'*) indicates the location of a pixel within a fine-tuning zone. Under such circumstances, the value of the local error is zero. By contrast, inconsistent overlapping of the two segments is indicated by the lack of a correlation between those segments. Calculation of the local refinement error between two segments is achieved based on the formula below:

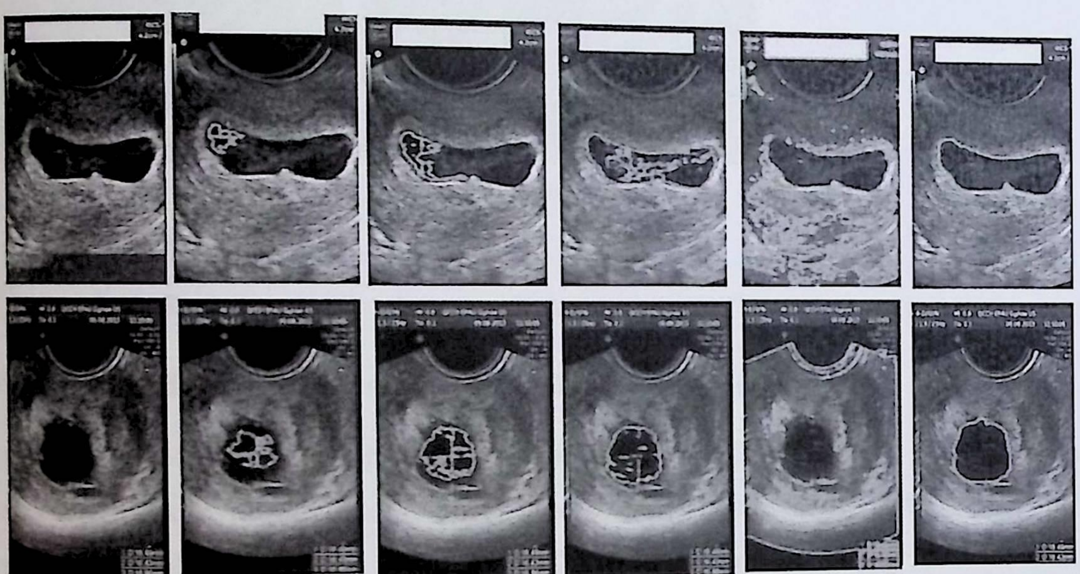
$$E(S_1, S_2, p_1) = \frac{|R(S_1,p_i)/R(S_2,p_i)|}{|R(S_1,p_i)|} \quad (6\text{-}6)$$

In the above, set difference is indicated by \, while the two segments are denoted by $S_1$ and $S_2$. For a particular pixel ($p_i$), the segments containing $p_i$ in *S1* and *S2* are considered and $R(S_1, p_i)$ indicates the pixel sets. When $S_1$ is a fine-tuning of $S_2$, $E(S_1, S_2, p_i)$ is equivalent to zero. However, this does not hold true when $S_2$ is a fine-tuning of $S_1$. Equation 6-6 is used to calculate the GCE between (*S*) and (*S'*). It is in the range 0-1 and a lower value is preferable.

$$GCE(S,S') = \frac{1}{n} \min\{\sum_i E(S,S',p_i \sum_i E(S',S,p_i)\} \quad (6\text{-}7)$$

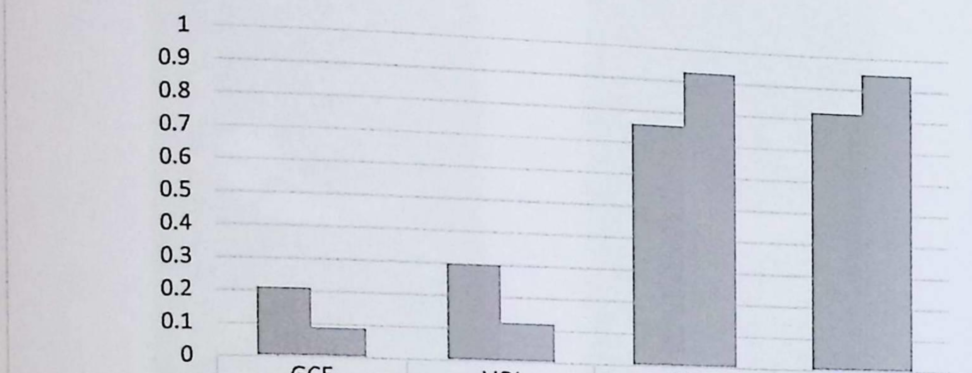
### A- Trainable RG vs Traditional RG

The visual evaluation of experimental results confirms that the trainable RG performs better than the traditional method. Figure 6-11 and Figure 6-12 shows the difference between the trainable RG and the proposed RG model using different image cases where we have applied the traditional RG with a different threshold to see the associated effect. We can clearly see that when we use a small threshold, we cannot catch the whole region due to the inhomogeneous region (under-segmentation problem). By contrast, when we increased the threshold, we will get more pixels from the ROI with less under-segmentation. When we increase it further, we will get the whole ROI with the over-segmentation problem. It is clear that the result of the trainable RG is better than the traditional one, as well as helping us to avoid estimating the threshold to get the entire ROI.



Original image | RG T= 0.12 | RG T= 0.14 | RG T= 0.18 | RG T= 0.2 | Trainable RG

***Figure 6-11: Region growing vs. trainable region growing model (pregnancy)***

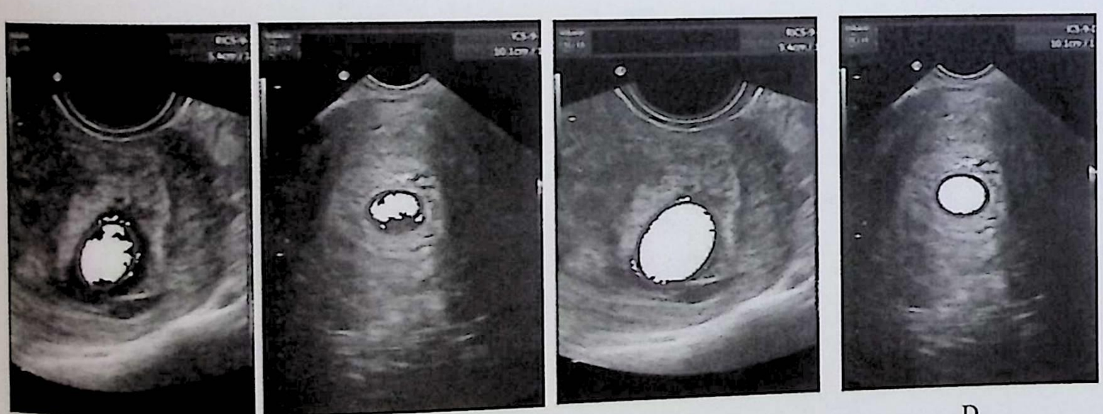



| | GCE | VOI | PRI | SSIM |
|---|---|---|---|---|
| RG | 0.21 | 0.3 | 0.76 | 0.823 |
| Trainable RG | 0.09 | 0.121 | 0.93 | 0.95 |

***Figure 6-13: Evaluation of the effects of the trainable RG vs. the RG***

**C- Manual vs. Automatic Measurements**

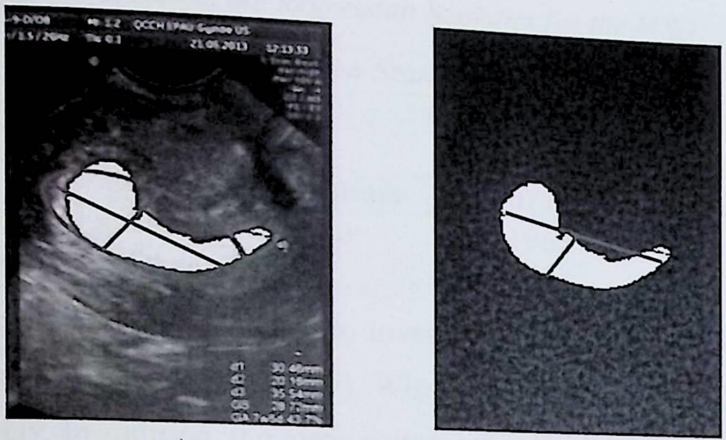
The aim of estimating the GS is to measure the MSD of the GS size. We have used the best-fitting ellipse for the segmented GS to estimate the MSD for the GS. Figure 6-14 shows the effect of the trainable RG on the MSD measurements.



A B C D

***Figure 6-14: The best fitting ellipse for both our previous algorithm and the trainable RG: A and B) represent the segmentation based on our trainable segmentation. C and D) represent the segmentation based on the RG trainable Segmentation.***

The fitting ellipse for irregular GS will negatively affect the MSD measurement, i.e., Figure 6-15-A shows that the ellipse will not produce an accurate MSD for the irregular GS, whilst the ellipse will not produce an accurate measurement for the MSD. Therefore, we have measured the MSD for the GS by calculating the miner and major dimensions for the segmented GS itself.



A B

*Figure 6-15: The best-fitting ellipse for irregular GS, A) measure the MSD based on the ellipse and B) measure the MSD using the minor and major dimensions for the object.*

Automatic MSD measurements were considered in relation to manual measurements to assess the degree to which the proposed system was effective. Figure 6-16 and Table 6-1 provides an indication of the performance. In view of the approximately 45° regression line, it can be concluded that the automatic measurements were remarkably close to the manual ones. It also seems reasonable to claim that there is no apparent systematic bias of the automatic measurements.

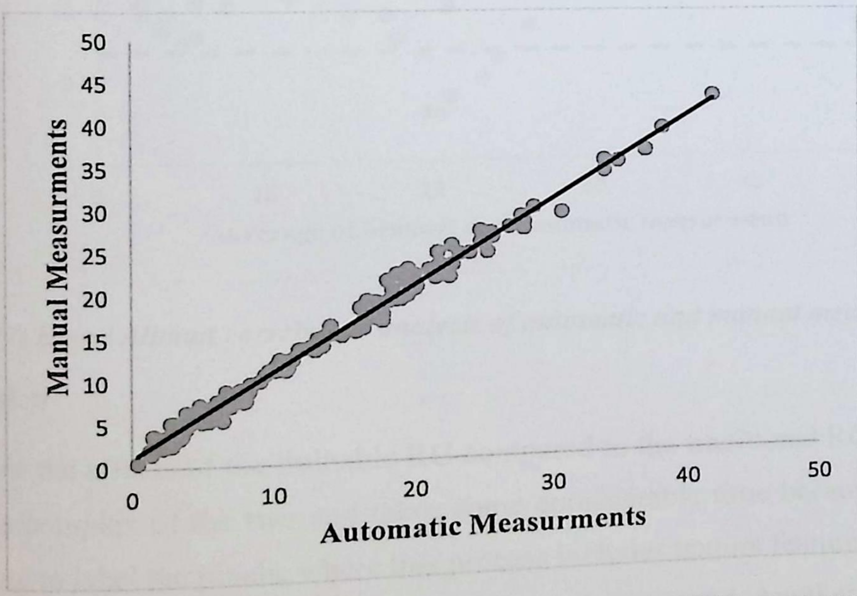



*Figure 6-16: Comparison between manual and automatic MSD measurements*

*Table 6-1: Shows the Regression Statistics for the MSD*

| *Regression Statistics* | |
|---|---|
| **Multiple R** | 0.995416 |
| **R Square** | 0.990853 |
| **Adjusted R Square** | 0.990795 |
| **Standard Error** | 0.871962 |
| **Observations** | 161 |

The Bland Altman procedure has been used to investigate the closeness between the manual and automatic measurements (see Figure 6-17). Where the Upper LOA and Lower LOA are --2.6 and 0.7 respectively. In addition, the mean square error (MSE) between manual and automatic measurement was calculated, where the MSE changed from 5.19 with previous algorithms to 2.20 when we use the trainable RG.

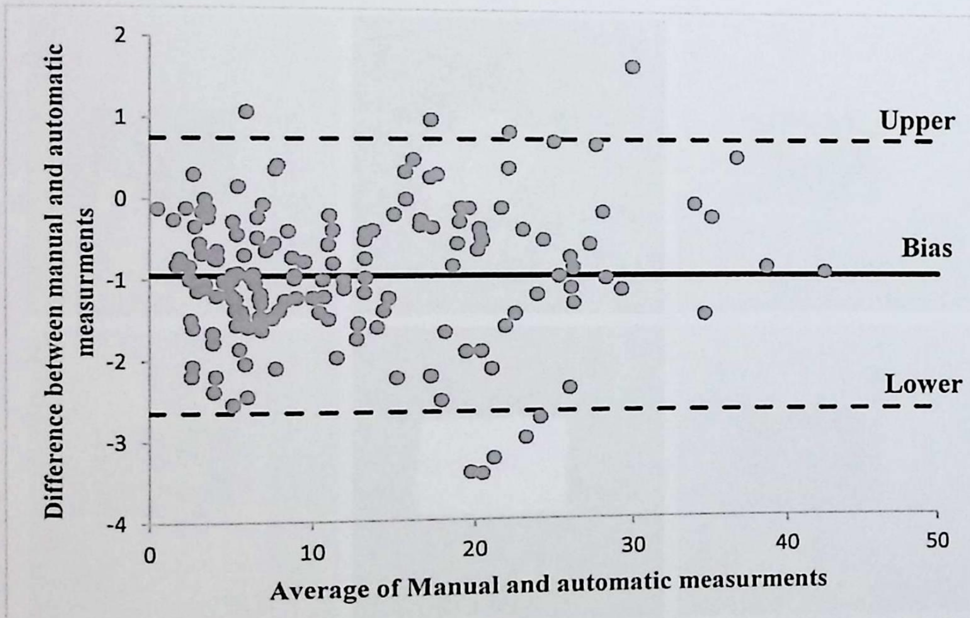



*Figure 6-17: Bland Altman correlation analysis of automatic and manual measurements*

### 6.2.4 Discussion

The results show the effects of the trainable RG compared to the traditional RG; the trainable RG is the more complex of the two and takes some considerable time because we used the machine learning to label the pixels, where this process includes texture feature extraction followed by ANN to label the pixels as part of the ROI or as background. Another issue we faced with this model was the need for a number of samples for the training phase, as well as the user selecting the samples needing experience to identify the ROI from other regions. Moreover, for different applications, we need to provide different training samples.

The results explained above show that the proposed model can work with the ultrasound images and will help to reduce the under-segmentation problem. Additionally, this model can help with the over-segmentation problem if we train the model carefully with different samples, including those of difficult cases.

To see the effect of the model in general, we tested it with different applications. In Figure 6-18, we tested the new approach on other difficult images, which here were multi-texture images. We have taken these cases just to show the effects of the trainable RG model on other applications. These images are extremely hard to segment due to the larger change in the pixels' grayscales. Figure 6-18 shows the results of both the traditional RG method with a threshold of 0.2 and the trainable RG.

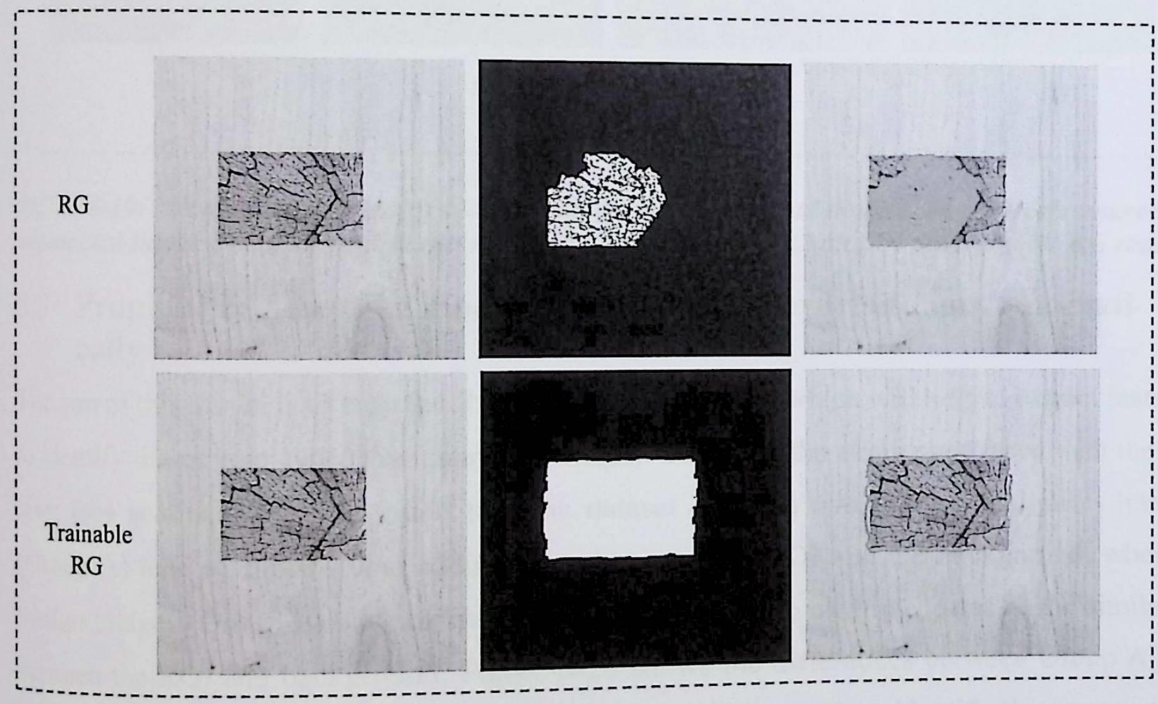



*Figure 6-18: The effective of the trainable RG with the multi-texture image*

Furthermore, the trainable RG has been applied to various challenging scenarios. Figure 6-19 shows samples from a dataset of microscopic images of blood samples (acute lymphoblastic leukaemia), which have taken from a public dataset. This figure includes four sections (A, B, C, D) for two sample. A) represents the original images, B) grayscale images C) segmented image based on traditional RG using a threshold of 0.2 and D) the trainable RG results.

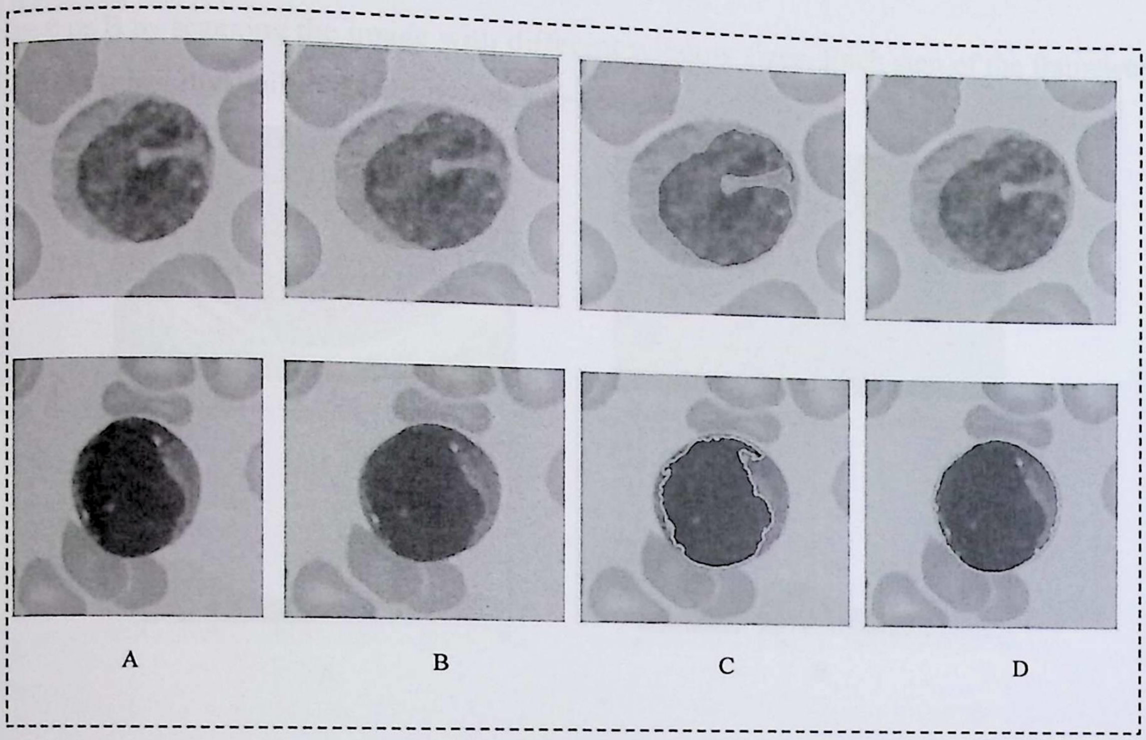



*Figure 6-19: Microscopic images of blood samples: A) the original images, B) grayscale images, C) segmented image based on traditional RG using a threshold of 0.2 and D) the trainable RG results.*

## 6.3 Proposal 2: Adaptive model to segment the ovarian mass automatically

The aim of this model is to crop the ROI of ovarian tumours, which will help to extract features to identify the tumour type (malignant or benign). Based on the challenges faced with the dataset (see section 6.1.2) we found that the dataset includes two groups: malignant images (Group A) tend to show a coarse difference between the ROI and the background, whereas benign images (Group B) have an indistinct ROI (there is no clear cyst) due to the similarity between the ROI and background. Figure 6-20 shows the differences between Group A and Group B. Distinguishing group A from group B can help to, first, identify the tumour type (benign or malignant) where the images with the clear cyst gives an indication that this case could be malignant and, second, use a suitable segmentation algorithm to detect the ROI and use it to extract the geometry and texture features to identify the tumour type. Figure 6-21 provides an overview of the system for the automatic identification of Type A and Type B, as well as detecting the ROI of the ovarian tumour from a static B-mode image. The framework consists of a sequence of stages starting from the stage 1 trainable model to identify group A from group B, then stage 2 to segment or detect the ROI using 1) the trainable model to segment the ovarian mass followed by filtering out non-ROIs based on texture features. This model has been used to segment image Group A. 2) Object detection model to detect the ovarian mass of image type B by scanning the image with different window sizes. Each step of the framework will be explained in detail in the following sub-sections.

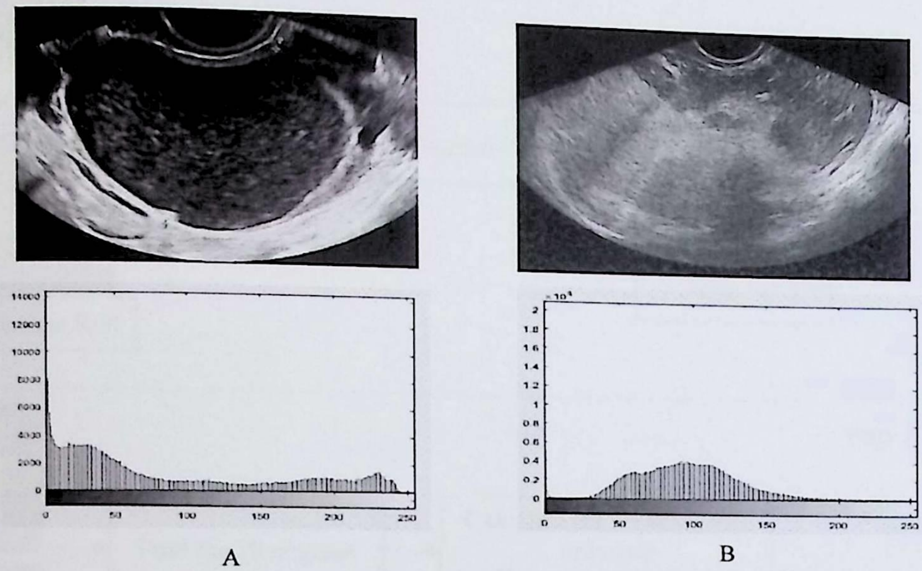


*Figure 6-20: Ovarian tumour samples A) Clear ROI (Group A) with its histogram and B) unclear ROI (Group B) with its histogram*

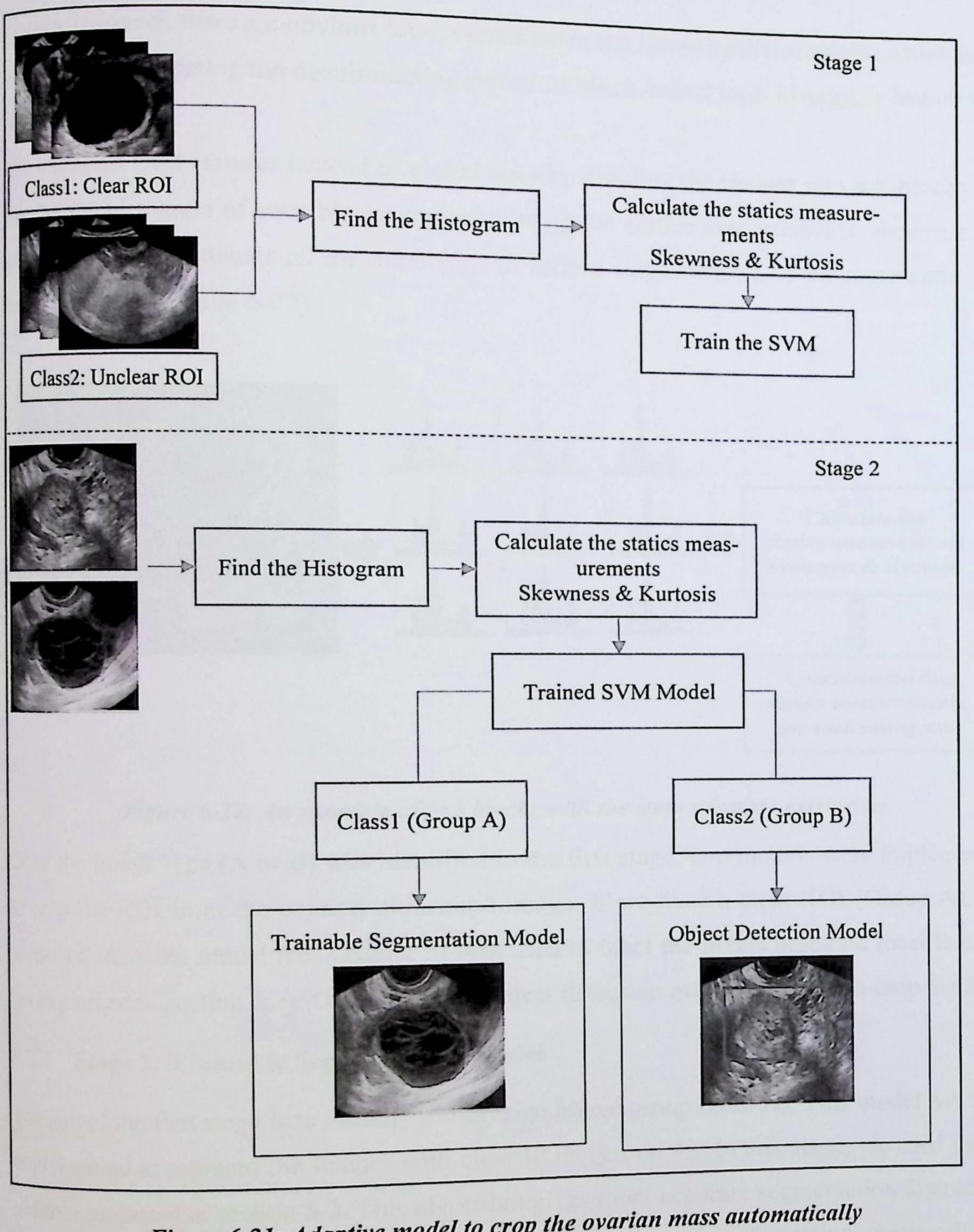



*Figure 6-21: Adaptive model to crop the ovarian mass automatically*

### 6.3.1 Stage1: Identify the cases with solid tumours based on the SVM classifier

As mentioned previously, the ovarian tumour dataset includes two groups of images (groups A and B). The aim of this stage to identify group A from group B. The process starts with the selection of a certain number of images for training purposes. For each image of either class, a two statics measurements, skewness and kurtosis, are extracted. Extracting feature vectors from the entire image may not be ideal for discriminating two ultrasound images of the different classes. However, there are obvious local variations in the intensity distributions, which is the motive for investigating the discriminating power of block-based local histogram feature vectors.

We extracted local features instead of global ones by dividing the images into sub-blocks, obtaining the histogram of each block, and calculating the statics measurements, skewness and kurtosis. Then concatenate all measurements to form a single feature vector representing the input image (see Figure 6-22).

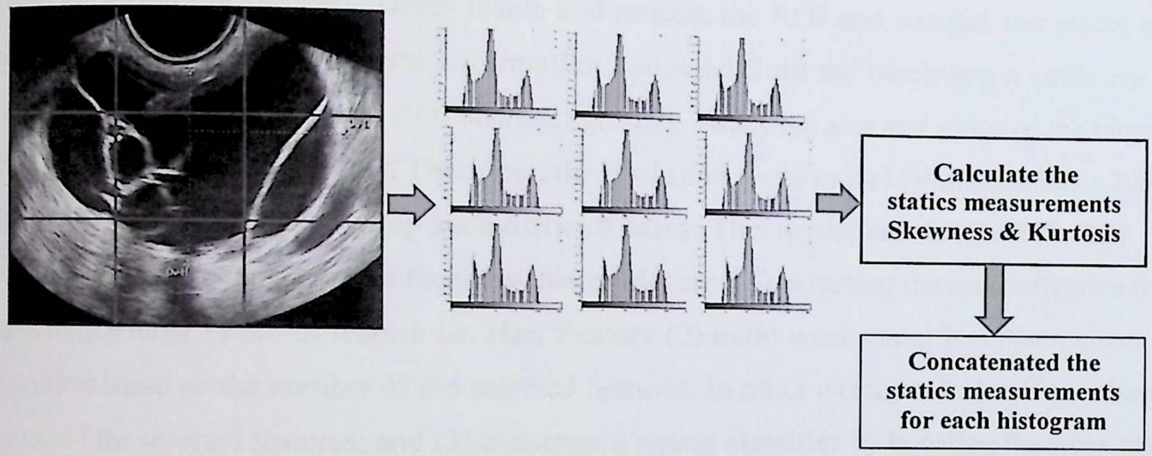



*Figure 6-22: An example of 3x3 blocks with the statics feature extraction*

Once the image type (A or B) was identified in the first stage, two models were implemented to crop the ROI from the ovarian ultrasound image. If we have a clear ROI (Group A), our proposed trainable model (see chapter 5) was used to label the pixels based on local features (as explained in section 5.3). Otherwise, the object detection model was used to crop the ROI.

### 6.3.2 Stage 2: Trainable Segmentation Model

The aim of the first stage is to identify the ovarian image group (A or B). This model has been implemented to segment the images with clear ROIs (group A). In this stage, we used the algorithm proposed in section 5-3. This algorithm gives more accurate segmentation than object detection because it will label each pixel based on the local texture information which will help to find the shape and the border of the ROI. As explained earlier, that model includes two stages: training and testing. In the training stage, we selected a number of images as training samples; from these images, we then selected a number of block samples from inside and outside the ovarian mass to build two classes, positive and negative. The HOG feature was used to describe each block. These features were feed to the ANN classifier to build a trained model. In the testing stage, the model scans the whole image with fixed-size blocks followed by extracting

HOG features. The trained model will label the blocks into binary values 0 or 1, to generate a binary image. Finally, we have used the cyst border estimation technique and watershed transform explained in sections 5.3.3 to estimate the border of the mass.

### 6.3.3 Stage 3: Object Detection Model

Images in group B show a great deal of similarity between the ROI and the background; as additionally, there is no clear border to detect the ROI. Therefore, our triable segmentation model based on fixed windows will be unable to locate the ROI due to, first, the overlap between the textures of the two classes inside and outside the ROI and, second, our model has classified a small window, and the information extracted from the block is not sufficient to capture the changes between the ROI and background. Third, the size and shape of the tumour changes from one case to another. Therefore, the Viola and Jones model (Viola and Jones 2001) (Cascade Model) was used to crop such difficult cases. This model has designed for face detection. The limitations that have faced by this model are: (1) selecting the most effective features from a large vector of feature i.e. Harr Feature (2) build weak classifiers, the number of classifiers based on the number of the selected features. In other word, each classifier is based on one of the selected features; and (3) construct a strong classifier by boosting the weak classifiers to.

One of the advantages of this model is that it scans the image with different window sizes, and this will help to capture the changes between the ROI and the background. On other hand, the main limitation facing this model with ultrasound images is the number of false positives produced by the algorithm in addition to the true positive. Figure 6-23 shows the true and false positives as a result of the Viola and Jones model as applied to an ultrasound image. Therefore, our objective is to enhance the Viola and Jones model to make it suitable for application to ultrasound images. We have improved it by add a filtering out stage as a post-segmentation technique to identify the correct ROI. Figure 6-24 shows the proposed system that consists of three stages: the training stage, testing stage and filtering-out stage.

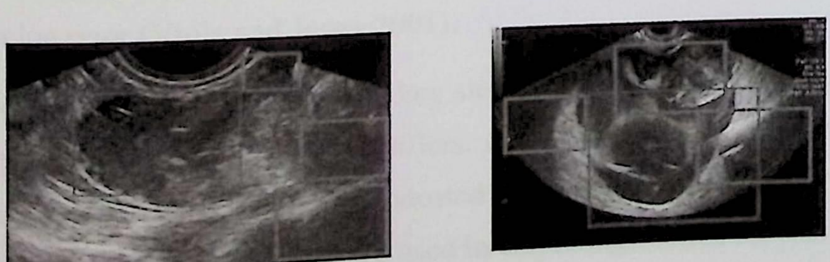


*Figure 6-23: The true positive and the false positives for two ovarian tumour cases.*

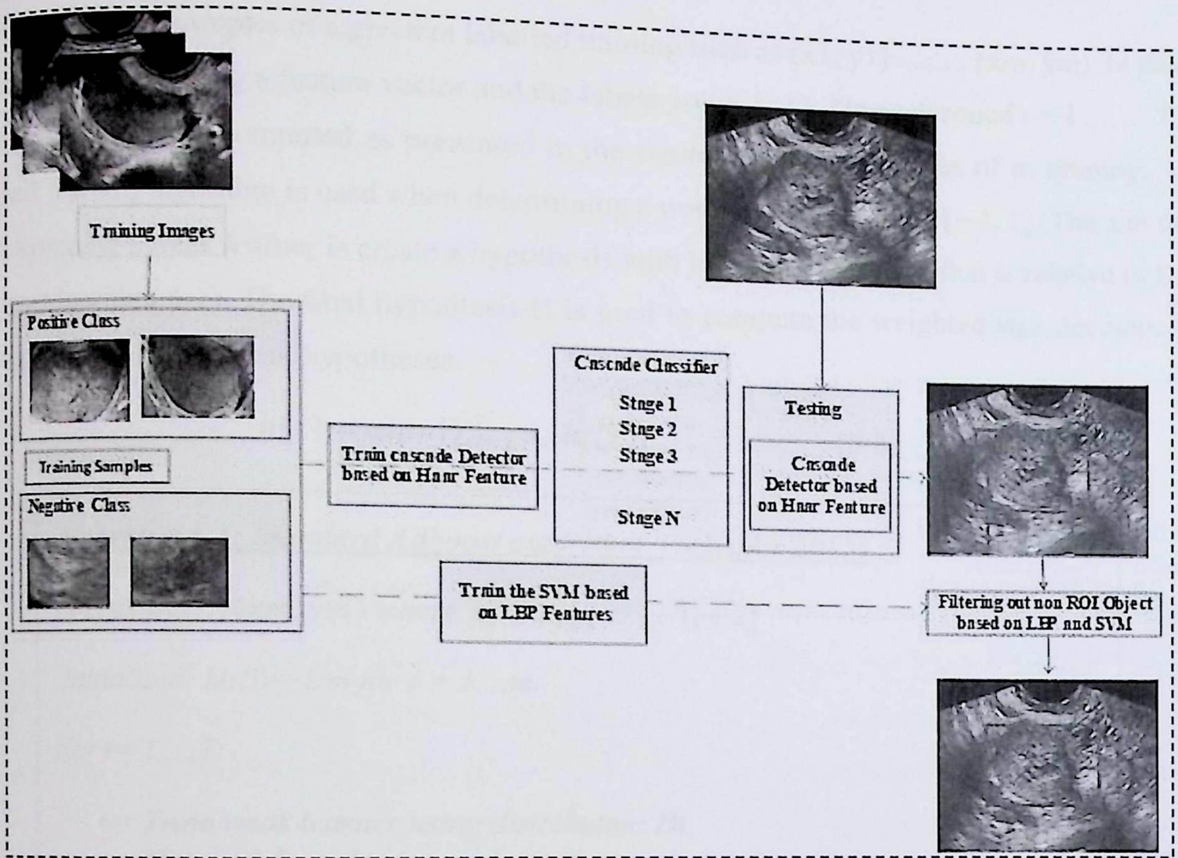



*Figure 6-24: Cascade Object Detector*

### A- Train the Object Detection Model

The aim of this stage is to build a classifier that can distinguish the true positive area from the false positive. In our proposed model, we have trained two models, as seen in Figure 5-24. First, the AdaBoost model was used to detect the true positive object and second SVM to filter out the non-ROIs produced by the first model. We have used the same training samples for both models. Those samples have been cropped manually.

The most commonly used Boosting algorithm is the AdaBoost, which has confirmed that combining weak classifiers helps in establishing a strong classifier. AdaBoost has proved its efficiency in combining simple statistical learners while reducing both the training error and vaguer generalization error (Viola and Jones 2001).

When using AdaBoost-based approach two key steps will be followed and they include constructing strong classifiers using weak classifiers. Another step will involve sequential combining of strong classifiers to develop a boosted classifier cascade. Multiple samples are cropped manually and the Harr Features are used in extracting vectors for each sample.

Here, there are examples of a given m labelled training such as (x1, y1)……., (xm, ym). In this case, x is representing a feature vector and the labels $y_i \in \{-1, 1\}$. On each round t = 1……. T, a distribution $D_t$ is computed as presented in the figure over the examples of m training. A weak learning algorithm is used when determining a weak hypothesis $x \to \{-1, 1\}$. The aim of determining a weak learner is create a hypothesis with low weighted error that is relative to $D_t$ (see Algorithm 6-1). The final hypothesis H is used to compute the weighted sign developed after combining the weak hypotheses.

$$h(x) = sign[\sum_{t=1}^{T} \alpha_t \, h_t(x)] \qquad (6\text{-}8)$$

***Algorithm 6-1: Standard Adboost algorithm*** (Schapire 2013)

$(x1, y1), \ldots\ldots, (xm, ym)$ $where\ x_i \in X, y_i \in \{-1, +1\}$

*Initialized: $D_1(i)$= 1/m for I = 1,..,m.*

*For t= 1,...,T:*

- *Train weak learner using distribution Dt.*
- *Get weak hypothesis* $x \to \{-1, 1\}$
- *Aim: select $h_t$ with low weighted error*

$$D_{t+1}(i) = \frac{D_t exp(-\alpha_t y_i h_t(x_i))}{z_t}$$

*where $Z_t$ is a normalization factor (chosen so that Dt+1 will be a distribution).*

*Output the final hypothesis:*

$$h(x) = sign\left[\sum_{t=1}^{T} \alpha_t \, h_t(x)\right]$$

**B- Test the object detection**

Viola-Jones proposed the basic principle of object detection algorithm. According to her principle, the detector is used to scan through the image many times with different sizes in every scan. Despite that an image is expected to have one or more object, it is apparent that negative results or non-ROI could still be achieved through excessive use of large amount of the evaluated sub-windows. This comprehension contributes to formulating different problems rather than evaluating the ROI and non-ROI should be discarded by the algorithm. This indicates that discarding a non-ROI is easier and faster than finding an ROI. With such assumption in mind, a detector with only one (strong) classifier appears to be ineffective because of constant evaluation time regardless of the input. This shows the need for a cascaded classifier.

The cascaded classifier consists of stages with each stage having a strong classifier. Each stage has a key role of determining whether a specific sub-window is absolutely a non-ROI or is an ROI. When there is a non-ROI sub-window in a given stage, the window is discarded immediately. On the other hand, when a sub-window is characterized as ROI, the sub-window is extended to the subsequent stage of cascaded classifier. This aligns with the assertion that the more stages that a given sub-window passes, the higher the chances of containing a ROI sub-window.

In a single stage classifier, accepting more false negative is the common problem when aiming at reducing the false positive rate. Nevertheless, during the initial stages of staged classifier, false positives are not the main problem. This is because the subsequent stages are projected to assist in solving the problem. Therefore, Viola-Jones provided clear recommendations to accept many false positives during the first stages. Therefore, during the final stages of staged classifier, minimal false negative cases are expected.

**C- Filtering out False Positive cases based on SVM**

To avoid false positives, the SVM classifier has been used to identify the correct ROI based on the LBP texture features, where the SVM was trained using the LBP and HOG features for both positive and negative training samples (see section 5-2-2). The trained SVM was used as a filter to reject the false positive cases and keep the true positive (ROI).

### 6.3.4 Proposal 2: Experimental Results

Automatic segmentation of the ROI is not a trivial problem and can be difficult task when the object of interest and the background share the same texture colour. In this work, we have proposed an adaptive model to crop the ROI. In general, to evaluate the segmentation methods, it is important to have the expert measurements and compare them with the automatic solutions to see the closeness and the correlation. The proposed model was tested on the ovarian tumour dataset described in section 1.4.2. Unfortunately, for that particular dataset, we have only the label class (benign and malignant) and there are no expert measurements as ground truths. Therefore, we have used the segmented image for another objective, which was to identify the ovarian tumour type based on texture features. The texture features should be extracted from the ROI to get the specific information related to the ROI only. There are a number of papers in the literature that identify the risk of the ovarian tumour based on the texture feature. We have selected the work of (Khazendar, et al. 2015) to compare to because they used the same dataset as that used in this thesis. In (Khazendar, et al. 2015), the authors cropped the ROI manually followed by feature extraction (histogram and LBP) to feed to the SVM classifier to diagnose the case. We have used the segmented image produced from our model and calculated the histogram of the greyscale image and LBP texture features to see the closeness between our model and the identification model proposed by (Khazendar, et al. 2015). Table 6-2 shows the diagnosis-based manual cropping and automatic cropping of our proposed method. We can clearly see that the proposed system was remarkably close to the manually cropped samples, demonstrating that the proposed system has cropped the most important region. Furthermore, Figure 6-25 and Figure 6-26 show segmented samples from Group A and B, respectively. We note that the proposed adaptive model can capture the exact border for cases that belong to group A. On other hand, images from group B have no clear border and therefore the proposed model estimates and ultimately crops the ROI.

***Table 6-2: Diagnostic performance of the Support Vector Machine on images processed using a Local Binary Pattern operator in the test group when using Radius R = 2***

| | **Average diagnosis for SVM without LBP** | | | **Average diagnosis for SVM & LBP** | | |
|---|---|---|---|---|---|---|
| **2x2 block image** | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| **ROI cropped Manually** | 68 [65-71] | 64 [61-0.68] | 66 [63-69] | 75 [73-77] | 72 [71-73] | 74 [73-75] |
| **ROI cropped by Our proposed model** | 66.45 [62-69] | 67.62 [64-71] | 67.7 [62-68] | 72.34 [70-74] | 75.23 [73-78] | 73.17 [71-78] |

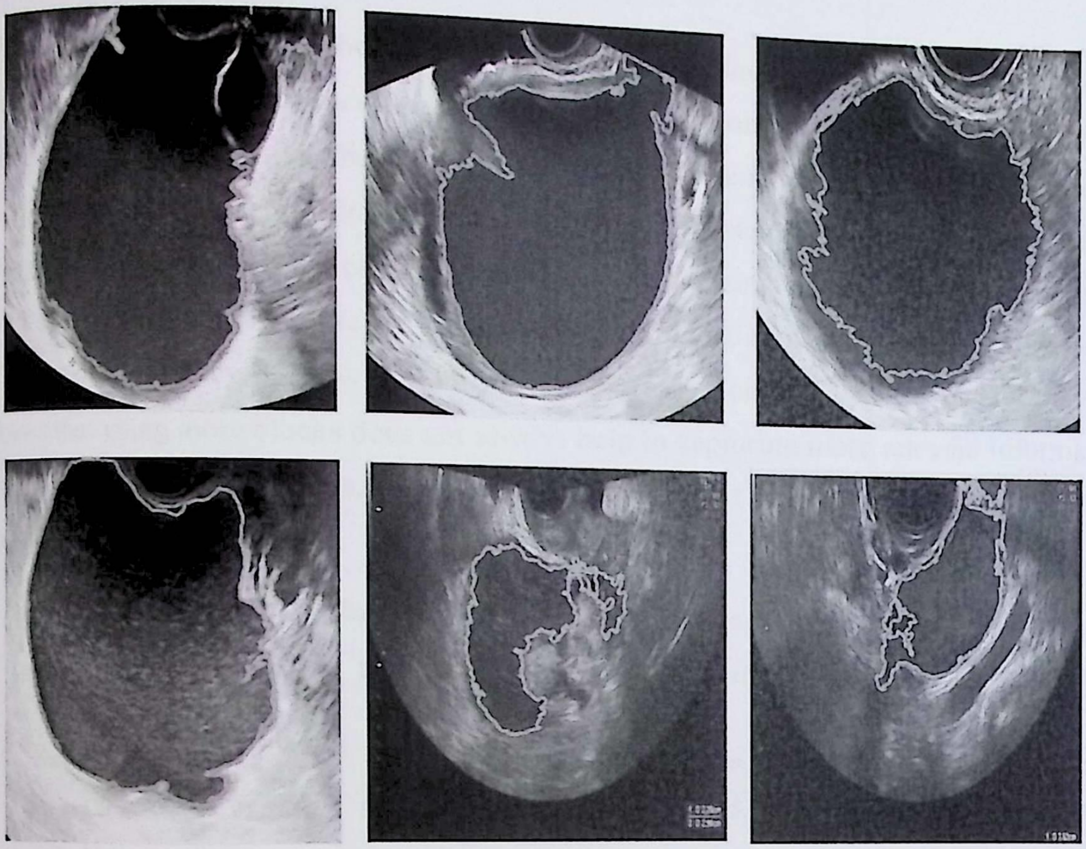


*Figure 6-25: Object detection for image Group A*

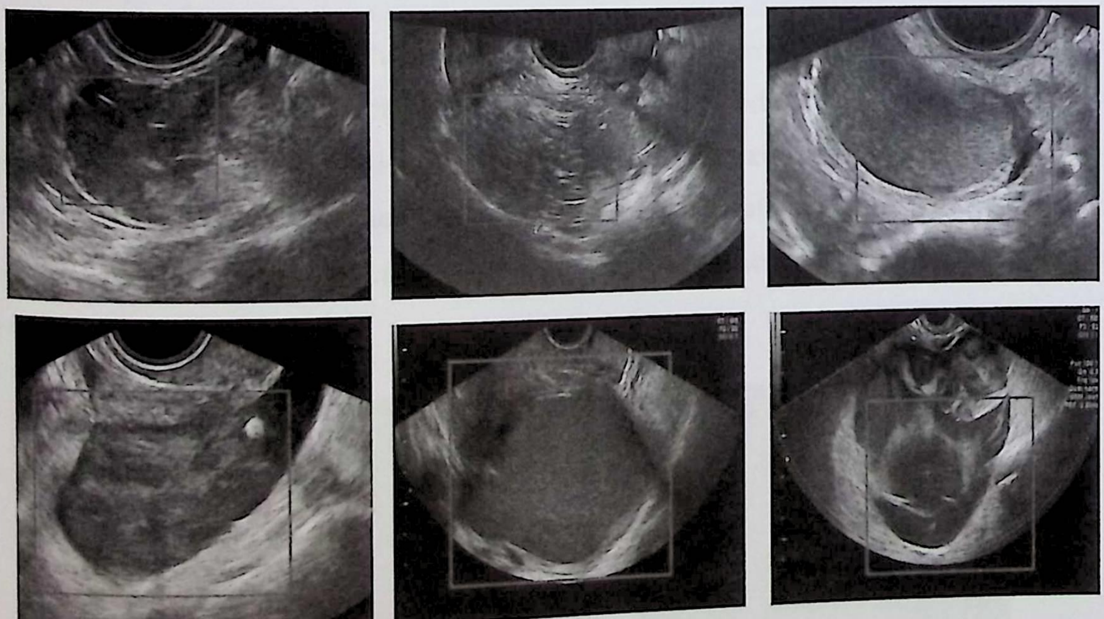


*Figure 6-26: Object detection for image Group B*

### 6.3.5 Discussion

We investigated the performance of eight different-sized block subdivisions. Table 6-3 demonstrates the effectiveness of local features over the global one. It shows the accuracy rate obtained by the SVM classifier for each of these eight different subdivision schemes. We notice significant improvements in terms of classification accuracy from 88% for using the whole image to over 94% for most of the blocking schemes attempted, and thereby confirmed that local features can capture more information relevant to distinguishing Group A from Group B. The optimal accuracy of 94% is achieved for the 5x5 blocking scheme, but the figure also shows that using more blocks does not always help in capturing more relevant information. In fact, when we have 8x8 blocks, the extracted features might become too localised, and thus the overall accuracy deteriorates.

*Table 6-3: The effectiveness of local features over global ones*

| | Whole Image | 2x2 | 3x3 | 4x4 | 5x5 | 6x6 | 7x7 | 8x8 |
|---|---|---|---|---|---|---|---|---|
| Accuracy | 88.8% | 88.3% | 87.3% | 89.4% | 94.01% | 91.3% | 87.1% | 82.54% |
| Class 1: Clear ROI (Group A) | 91.26% | 89.18% | 88.75% | 90.41% | 95.73% | 92.44% | 87.9 | 83.9% |
| Class 2: Unclear ROI (Group B) | 86.38% | 87.5% | 85.89% | 88.48% | 92.4% | 90.23% | 86.4 | 81.1% |

There are number of limitations that we have faced with the proposed model, i.e., in stage 1 we have some cases which had a clear border but were classified as Group B (see Figure 6-27). The main reason for getting a false positive is due to the texture of the ROI for those cases having a texture similar to the background. Therefore, the statics features (skewness and kurtosis) failed to identify it as belonging to group A, as well there not being enough samples to cover all cases.

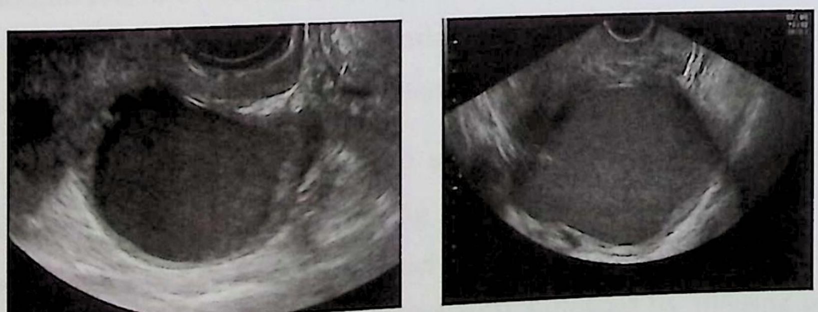


*Figure 6-27: Examples of the false positive for stage 1*

In addition, there are some cases that show a high similarity between the ROI and surrounding area (see Figure 6-28). These cases are exceedingly difficult, even for an expert, because there is no clear change that can help to detect the ROI.

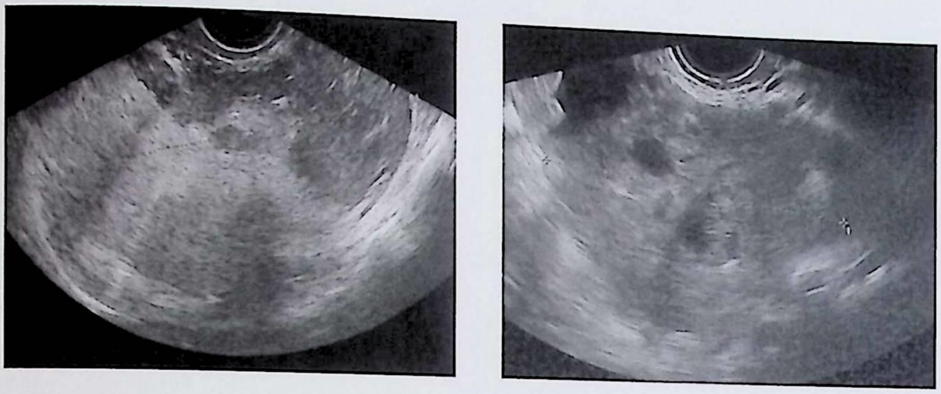


***Figure 6-28: Examples of difficult cases.***

Finally, another limitation facing the proposed system is the filtering out stages incorrectly identified the ROI for filtering, which we strongly believe is due to with the limited number of samples available, and therefore that we cannot cover all cases.

## 6.4 Summary

This chapter describes two approaches to enhance the trainable segmentation. First, for the trainable RG approach to enhance segmentation quality, we argued that RG based on intensity or homogeneity was feasible. Failure of the system was either due to the threshold being unsuitable for some of the images, or due to poor image quality. The proposed method is unique in comparison to previously reported methods because it performs a complex image segmentation process based on texture feature using an artificial neural network. This process was helpful in image regions with poor image quality. Comparison between the trainable RG approach and the traditional RG showed excellent agreement. In addition, accurate segmentation in the proposed and the new approach are very similar. Less agreement was found between the trainable RG and manual tracing for areas calculated from poor quality images.

Second, we have proposed an adaptive model to segment the ovarian tumour images. The framework consists of a sequence of steps starting from stage 1, trainable model to identify Group A from B, then stage 2 to segment and detect the ROI. Stage 2 includes two models: 1) trainable model to label the blocks and identify the ROI followed by filtering out non-ROI based on texture features. The proposed trainable model has used to segment image type A. 2) Object detection model to detect the ROIs of image type B by scanning the image with different window sizes. This framework helps to crop the ROI for difficult cases. Experimental results demonstrate the effectiveness of the proposal in producing accurate detection comparable with those achieved by manual segmentation.

# Chapter 7. Texture-based Analysis of Ultrasound Images in Gynaecology

In the previous chapters we developed a multi-level trainable schemes to segment ROI in US scans of the ovary, with particular interest in the gestation sac for pregnancy assessment and tumour masses for cancer diagnostics. Pregnancy assessment is usually based on determining the MSD of the GS, and accordingly we devised an automatic algorithm to determine the MSD from the segmented GS and used the medical expert criteria to identify miscarriage cases from PUVs. This was particularly successful in that there was a high correlation with the decisions made by the medics using these US scan images. However, comparison of our scheme's decisions with the ground truth provided by the hospital revealed that the MSD-based initial diagnosis agreement with the ultimate diagnosis was quite high when the computed MSD sits comfortably within the medically practiced criteria. However, as the computed MSD deviate from the 'comfort' zone, disagreements with the ground truth started to increase. This raised the new challenge of the search for other image-based indicators beyond the MSD that could be used with some machine learning schemes to complement the MSD-based decision. This chapter was devoted to such an attempt, and we shall demonstrate that machine learning using texture-based features can result in improved identification of miscarriage cases from the early stage US images.

As mentioned previously, the objective of the previous chapters was to extract the object of interest from the ovarian tumour images followed by measure feature, i.e. geometry features (existing expert measurement) and texture features (i.e., LBP). The texture features can be used as a new sign to support the expert decision. Based on our investigation and the literature review, we found that speckle noise is one of the critical factors that can affect the accuracy of diagnosis. Therefore, this chapter is concerned with a proposed approach based on a trainable model for ultrasound speckle suppression. This approach includes two stages, 1) trainable model to identify the complexity of the speckle noise, and 2) filtering out the speckle noise.

The rest of this chapter is organised as follows: in section 7.1, we discuss our motivation for going beyond the MSD-based identification of miscarriage cases in early pregnancy as well as discussing the effect of the speckle noise on the diagnosis of ovarian tumours. Section 7.2

presents our approach to automatic discrimination of miscarriage cases using texture features. In section 7.3, we propose a model to enhance ovarian ultrasound images. Section 7.4 summarises the work carried out in this chapter.

## 7.1 Problem statement

### 7.1.1 Miscarriage identification with/without MSD parameters.

Once a pregnancy test produces a positive result, a scan is conducted, especially if the pregnant woman reports pain in the abdomen or there is bleeding, as well as when there is a history of any miscarriage. The results of such an ultrasound scan are used to inform the preliminary diagnosis. Gynaecologists use the MSD of the GS, manually measured from a US scan image of the ovary, as an effective benchmark for distinguishing MC from PUV. Different studies have considered that miscarriage should be declared based on different cut-off values for mean gestational sac diameter (MSD) within the range of 16-25 mm (Bourne 2016), where the American College of Radiology guidelines used a cut-off value of 16 mm as the threshold on which to make a decision. By contrast, the Royal College of Obstetricians and Gynaecologists of the United Kingdom uses 25 mm.

In the first scan (present scan), the Gynaecologist will use a cut-off value to identify the case as a miscarriage or PUV. In case of miscarriage, no additional scan is carried out, but if a miscarriage is not confirmed, testing is repeated after fourteen days, including an ultrasound scan and blood test (ultimate diagnosis). In our approach, we will use the two tests (present scan and ultimate diagnosis) as the ground truth to investigate the effects of the texture features.

The present study employed a total of 184 ultrasound images. Based on MSD (present scan), 157 images were labelled as PUV and 27 as miscarriage. Table 7-1 indicates that, in the ultimate diagnosis, numerous cases initially classified as PUV were subsequently reclassified as miscarriage, whereas a few cases lacked diagnosis labels (9 images; lost to follow-up, 44 images without description) where all these cases were excluded from the dataset.

*Table 7-1: The number of images in each diagnosis*

| Class name | Diagnosis on presented scan | Ultimate diagnosis |
|---|---|---|
| Not Miscarriage | 157 | 30 |
| Miscarriage | 27 | 101 |
| Total | 184 | 131 |

To outline the rate of false positives in miscarriage diagnosis, as well as to enable a more comprehensive diagnosis, this study sought to introduce new signs. This chapter will focus on two points to identify miscarriage cases in the early stages of pregnancy based on texture information. Firstly, the texture features extracted from ultrasound images of GS could offer values for diagnosis of miscarriage cases besides MSD. Secondly, can such feature vectors help in confirming MC cases (MSD > 25mm), i.e., offering any added value. Two reference points of diagnosis ground truth have been used: (a) diagnosis on the presented scan, and (b) ultimate diagnose.

The texture of ultrasound images has been considered in various types of research in order to characterise the echo-texture of B-mode images quantitatively. The principles of analysing the texture of B-mode ultrasound images are if disease procedures affect the structure of the tissue, the tissue will reflect a change in ultrasound signal, which will give different valued texture features to normal tissue (Morris 1988). Based on this, it is expected that texture features derived from cancerous and normal tissues will be different. Therefore, this work will focus on using ultrasound pregnancy images to determine five distinct features, namely, uniform local binary pattern (ULBP), fractal dimension (FD), a histogram of orientation gradient (HOG), and the Gary-Level Co-Occurrence Matrix (GLCM), enabling the extraction and collation of a number of features.

### 7.1.2 The impact of the speckle noise on the ovarian tumour identification

The complexity to identifying speckle noise (speckle noise level) is an important section in the selecting of the type of filtering which is necessary to enhance the ultrasound image, i.e., the image with the brighter region after applied speckle noise will present a magnified view, and greater random variations in pixel intensity are observed. On the other hand, when this type of noise is present in a darker area in the image, the random variation observed will be not high compared to the same observation in the brighter regions. Therefore, this kind of noise is signal dependent and distorts the image in large magnitude. In this chapter, we propose a new approach based on the trainable model to identify the speckle noise levels with the subsequent application of a suitable filter.

Figure 7-1 shows three types of thee ovarian ultrasound image. First, the ROI is clear, and this type of image has reduced levels of speckle noise. Second, the ROI has a texture representing reflection off the solid tissues. The third represents images that are a mix of the first and second types.


*Figure 7-1: Ovarian tumour samples illustrating the speckle noise level: A) Low-level speckle noise (not solid); (B) high level of speckle noise (solid texture); (C) combination of the two levels (solid and not solid)*

De-noising and enhancing the ultrasound images of an ovarian tumour helps 1) experts to improve their diagnosis by highlighting the anatomical information 2) to enhance the texture feature to reduce the overlap between the benign and malignant class. The general requirements for removing artefacts in ultrasound images are to suppress speckle noise as much as possible without affecting the anatomical information. We have proposed two models to reduce the speckle noise based on both global and local information.

## 7.2 Machine Learning for Early Miscarriage Identification

A model for automatic identification of miscarriage cases is proposed. The proposed CAD technique for such identification has been shown in Figure 7-2 in form of a block diagram. The CAD technique mainly consists of a classification system whose function is to predict the class label (miscarriage or PUV) of a test image based on a neural network classifier determined by a learning system.

As mentioned previously, each image is presented in two planes (sagittal and transverse). The framework consists of a sequence of steps starting from ROI detection. This is followed by a feature extraction technique to extract the features of the ROI for each plane. In the classification stage, we trained the system based on two datasets (sagittal and transverse). Therefore, we have two neural networks, one for the sagittal and the other for the transverse plane. The two neural networks will be used to classify the testing image.

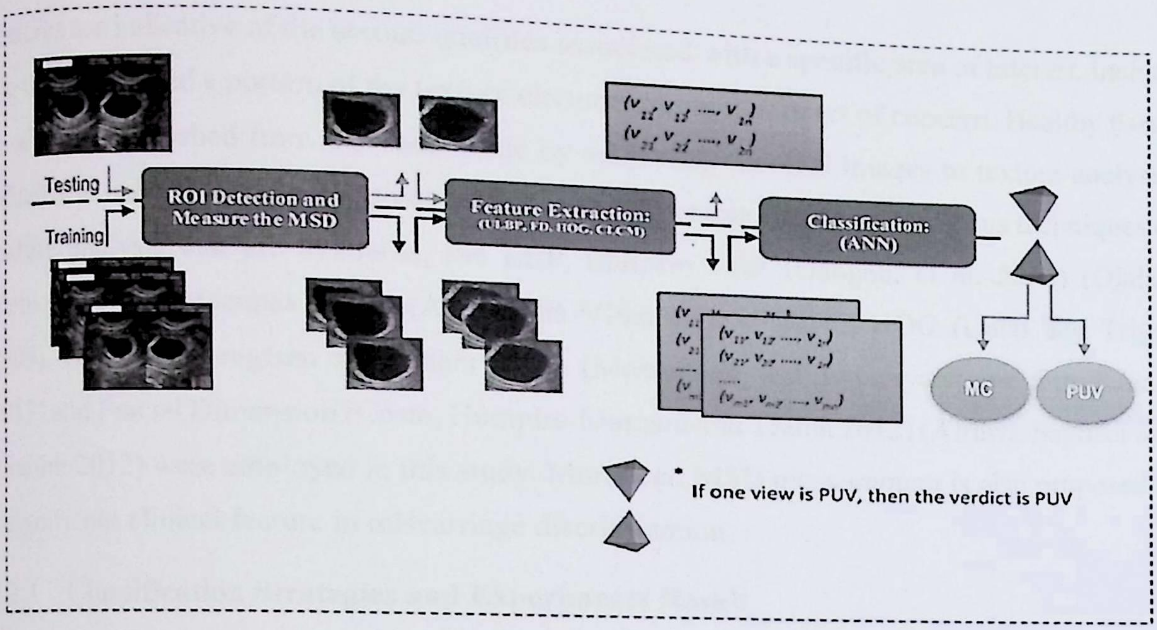



*Figure 7-2: Texture-based analysis for miscarriage identification in early pregnancy*

### 7.2.1 ROI Detection

In the previous chapters, we have proposed methods to segment the GS and extract the MSD from the segmented GS. This method has helped us in two ways, the first using the location of the segmented GS to crop the ROI which includes the GS with the edge and a small part of the surrounding area, whilst the second uses the measured MSD as an indicator to confirm miscarriage cases that have MSD > 25.

### 7.2.2 Feature extraction

The clinical description of data in terms of image texture had been done originally. In our example, the PUV images are expected to have a homogeneous distribution while the MC images, on the other hand, are expected to have an assorted grey-level distribution. In mathematics, in order to improve the final results, a blind set of texture-based features are selected from several reference articles. The quantitative attribute of image texture indicates how smooth the surface geometry is and how the object, as visualized in the image, is structured. In the context of image processing, texture refers to the manner in which the pixel grey levels are spatially distributed in an image or a particular target area.

The analysis and interpretation of medical images depends greatly on texture. As mentioned above, texture is a function of the spatial arrangement and diversity of pixel grey values (intensities) of B-mode images. Texture features can usually be determined through a number of approaches. Outlining a series of texture features is the main goal of texture analysis. These features are indicative of the texture qualities associated with a specific area of interest, including the margin and a portion of the texture circumscribing the object of concern. Healthy tissue is often distinguished from diseased tissue by subjecting medical images to texture analysis, which also helps to segment particular anatomical structures. Among the various techniques of texture analysis that are available, the LBP, uniform LBP (Gangeh, et al. 2016) (Ojalaa, Pietikainen and Maenpaa 2002) (Aramendía-Vidaurreta V 2016), HOG (Dalal and Triggs 2005), GLCM (Albregtsen and others 2008) (Mohanaiah, Sathyanarayana and GuruKumar 2013) and Fractal Dimension (Costa, Humpire-Mamani and Traina 2012) (Alrawi, Sagheer and Ibrahim 2012) were employed in this study. Moreover, MSD measurement is also proposed as a significant clinical feature in miscarriage discrimination.

### 7.2.3 Classification Strategies and Experiments Result

As explained previously, the dataset was labelled based on two reference points of diagnosis ground truth: (a) diagnosis on the presented scan, and (b) ultimate diagnosis. The Backpropagation ANN has been used as a classifier to identify the Miscarriage cases from the PUV for both present and ultimate diagnoses. In the following points we are going to explain the experiments that have been tried to test the proposed model.

a) **Diagnosis on presented scan:** three test options have been tried to test the features in this section are:

- Option One – Using the provided testing set: use the whole of the first batch as *training* and the whole of the second batch as *testing* examples. (Problem: class imbalanced)
- Option Two – Using samples of balanced classes: first merge the two batches and then take a stratified random sample of balanced classes (25 PUV and 25 MC). Then use a percentage split for each sample (60% as training (50% training, 10% validation) and 40% as testing). Repeat the sampling and the classification process 30 times.
- Option Three: Taking the first batch as the data source for training examples and the second batch as the data source for testing examples, sampling for training and testing examples of balanced classes. Repeat the sampling and the classification process 30 times

Figure 7-3 shows the measured averages of overall accuracy, sensitivity and specificity using the first option. We can see that the LBP texture feature has some diagnostic value in separating MC from PUV. By contrast, HOG, GLCM and FD are much worse than LBP. Low sensitivity and high specificity for other features are very much due to the small number of MC and a large number of PUV cases in the training set.

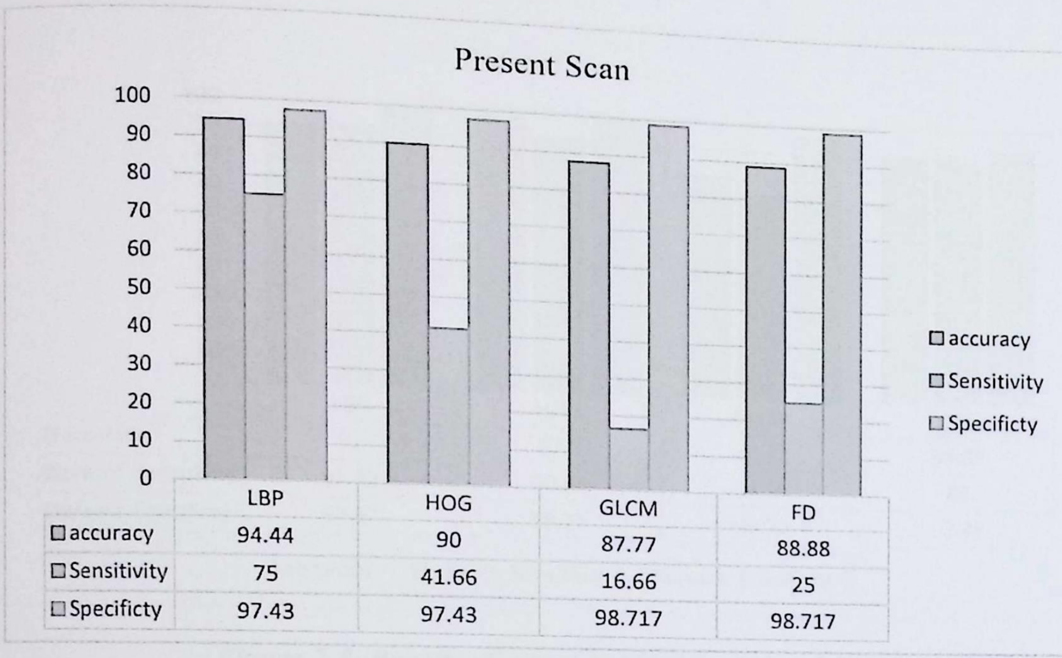



***Figure 7-3: Results of experiment using option one***

Figure 7-4 and Figure 7-5 show the results of the experiment using options two and three. We note that all textural features have some diagnostic value, but that HOG and LBP consistently show better performance than the other two. We strongly believe that the circularity represented by the HOG feature of the GS may indicate which class the GS should belong to. On the other hand, LBP representations of spots, edges, and border regularity may themselves indicate which class the GS should belong to.

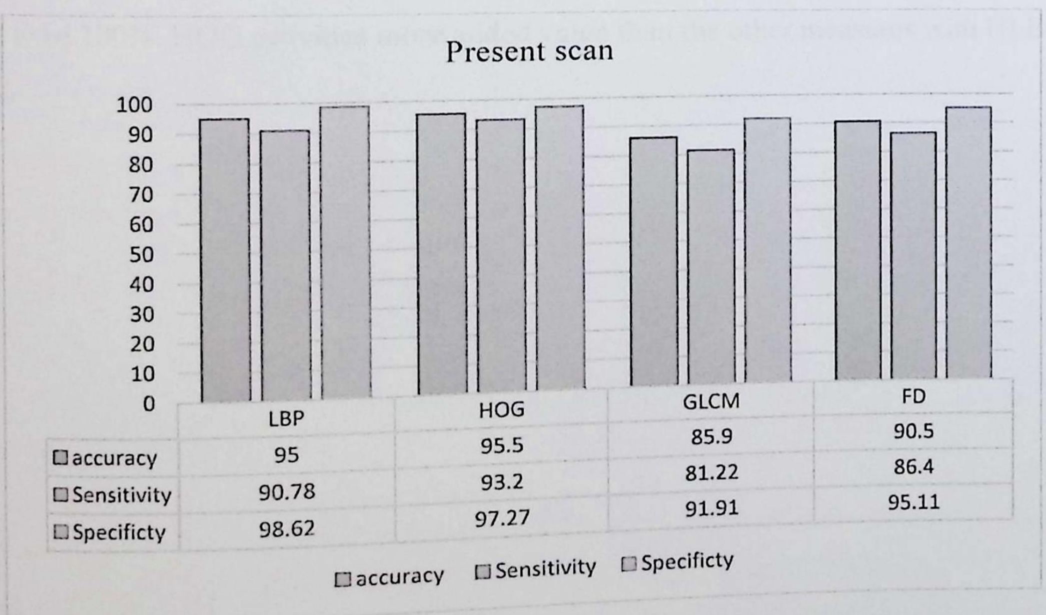



***Figure 7-4: Results of experiment using option two***

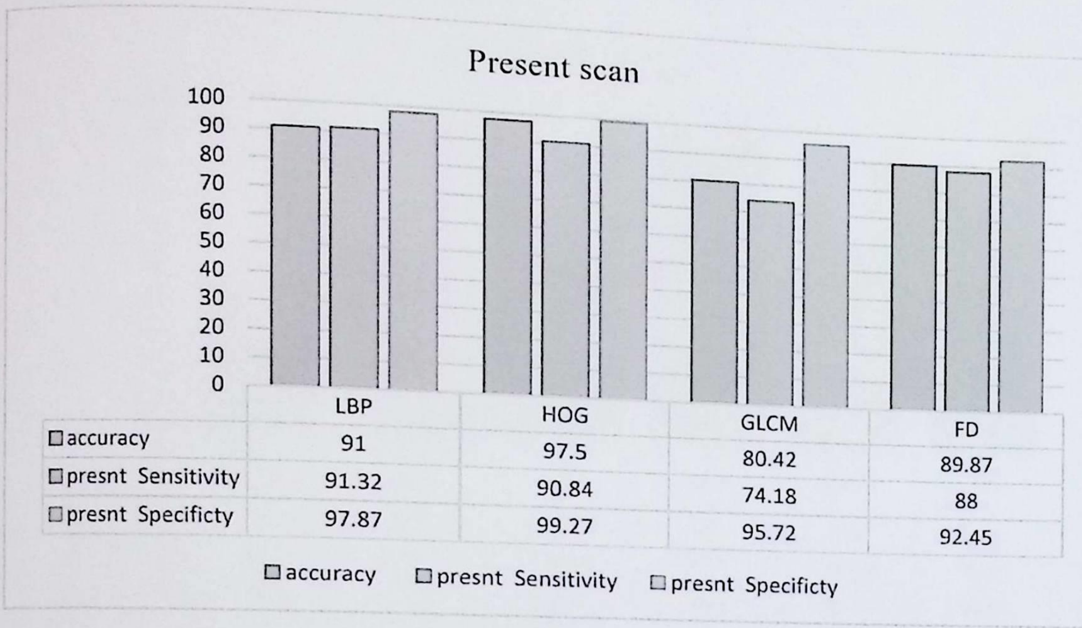
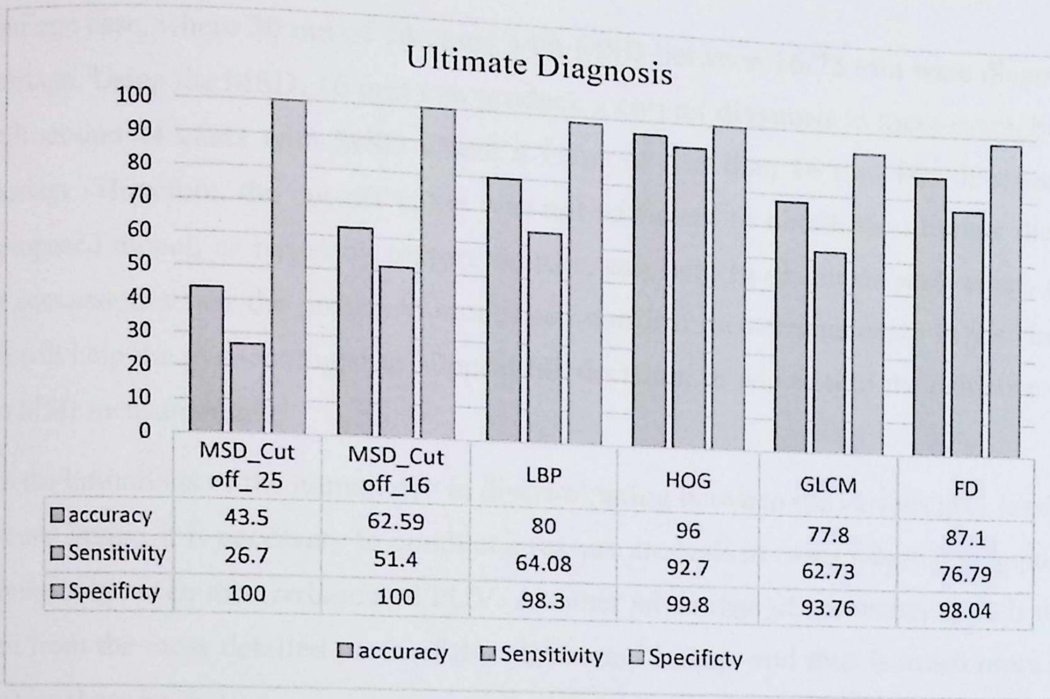


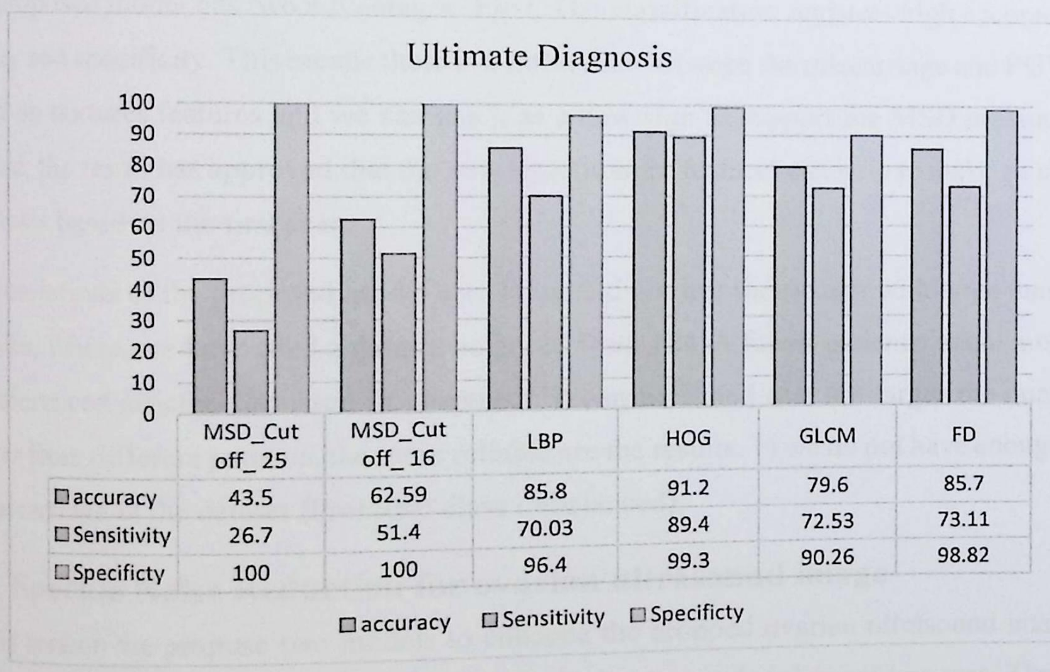

*Figure 7-5: Results of experiment using option three*

(b) **Ultimate diagnose:** two test options have been employed to test the features in this section. Option two and three, as explained in section (a), have used to evaluate the features. We cannot use option one because some cases lacked diagnosis labels in batch 1. The averages of overall accuracy, sensitivity and specificity have been measured. The figures below show the results of the experiment using option two (see Figure 7-6) and three (see Figure 7-7). It may be noticed that many PUV cases become MC when using both cut-off MSDs (16 -25). All textural features indicate added value in identifying the MC cases, although specificity cannot be guaranteed to be 100%. HOG provides more added value than the other measures with ULBP tails behind.




| | MSD_Cut off_25 | MSD_Cut off_16 | LBP | HOG | GLCM | FD |
|---|---|---|---|---|---|---|
| accuracy | 43.5 | 62.59 | 80 | 96 | 77.8 | 87.1 |
| Sensitivity | 26.7 | 51.4 | 64.08 | 92.7 | 62.73 | 76.79 |
| Specificty | 100 | 100 | 98.3 | 99.8 | 93.76 | 98.04 |

***Figure 7-6: Results of experiment using option two***




| | MSD_Cut off_25 | MSD_Cut off_16 | LBP | HOG | GLCM | FD |
|---|---|---|---|---|---|---|
| accuracy | 43.5 | 62.59 | 85.8 | 91.2 | 79.6 | 85.7 |
| Sensitivity | 26.7 | 51.4 | 70.03 | 89.4 | 72.53 | 73.11 |
| Specificty | 100 | 100 | 96.4 | 99.3 | 90.26 | 98.82 |

***Figure 7-7: Results of experiment using option three***

### 7.2.4 Discussion

The diagnosis based on the MSD in the present scan has a different cut-off value, i.e., in the USA, 16 mm is the cut-off value used to identify miscarriage cases, whilst in the UK 25 mm is the current standard. Based on the ground truth for the ultimate diagnosis provided by the hospital, we found that there were 74 cases them MSD lees than 25m and they diagnosis as a

Miscarriage case, where 30 out of 74 cases with MSD between 16-25 mm were diagnosed as miscarriage. Using the MSD, 16 mm can produce a correct diagnosis in these cases, but there are still around 44 cases with MSD where a value of less than 16 mm was diagnosed as a miscarriage. Therefore, the cut-off value was not sufficient to obtain the ultimate diagnosis. The proposed model, as based on texture feature, can help to diagnosis such cases, and the result demonstrates that the proposed model can confirm miscarriage cases in the first scan, which will help the gynaecologist to support his decision, in addition to the evidence offered by the MSD measurements.

Due to the limitations of the human eye in discriminating between the various grey levels in an ultrasound image, it is necessary to conduct a texture analysis in cases where it is important to discriminate between miscarriage and PUV. Another advantage of texture analysis is that it is created from the most detailed parts of the ultrasound image, and thus is much more critical than a visual analysis.

The proposed model has two advantages. First, The classification registers high accuracy, sensitivity and specificity. This means there is a difference between the miscarriage and PUV cases based on textures features and we can use it as a new sign to support the MSD measurement. Second, the result has approved that the new sign (texture feature) can help to make an ultimate diagnosis based on the first scan.

The limitations of the proposed model are: 1) we did not test the system with large number of samples. where, we have used a dataset acquired from 184 different patients, equal more than the referenced articles discussed in chapter 3, it can be stated that the larger the number of images from different patients, the more reliable are the results. 1) we do not have enough cases of miscarriage in the dataset (Problem: class imbalanced).

## 7.3 Speckle Noise Reduction for ovarian ultrasound image

In this section we propose two models to enhance the cropped ovarian ultrasound image and reduce the overlap between the texture feature for benign and malignant tumours. These two models are based on global and local information. In this section, the two models will be described in detail.

Figure 7-8 shows the model proposed for the identification of ovarian tumour type as based on texture features. The first step involves cropping the ROI and, in this step, the adaptive model proposed in chapter 5 has been used. Then, in the pre-processing step, an adaptive model was used to enhance the ROI, as per our focus in this framework. The enhanced ROI was used as input to the feature extraction technique, followed by the classifier to identify malignant tumours from benign. Our objective in this framework was to propose an adaptive trainable model to enhance the ROI. Therefore, the following sections will focus on the proposed speckle noise reduction model and evaluate it by considering the final result of the SVM.

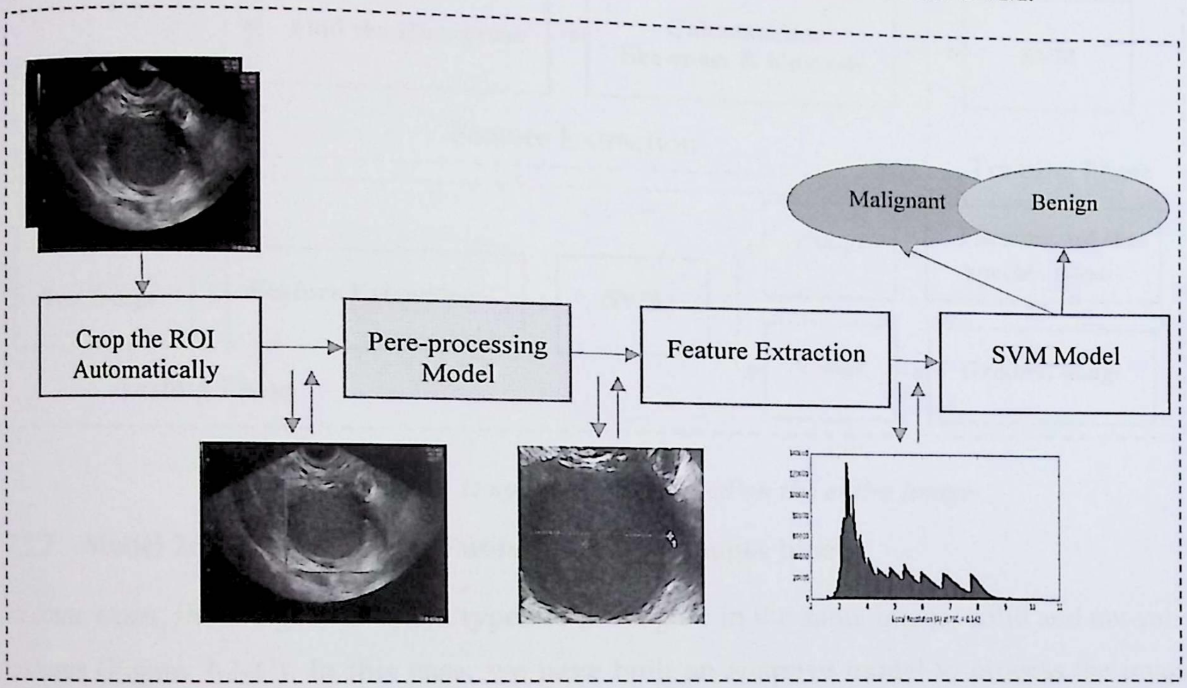



*Figure 7-8: Automatic identification of ovarian tumours based on texture features*

### 7.3.1 Model 1: speckle noise reduction based on the whole image

As mentioned previously, the complexity associated with identification of speckle noise is an essential factor in the selection of the type of filter necessary to enhance the ultrasound image. The proposed model includes two main stages: first, identify the complexity of the speckle noise; second, apply a filter to reduce the speckle noise (See Figure 7-9).

The process starts by selecting a number of images from the two classes (solid texture or otherwise) for the training stage. The statics measurements, which are skewness and kurtosis, were extracted from the histogram of the ultrasound image. Then, we feed these features into the SVM classifier to build a trained model. The testing phase includes processing the image by calculating the skewness and kurtosis from the histogram. The trained SVM model will be used to label the testing image as class 1 (black ROI) or class 2 (it has a solid object). We have not used a filter for the images labelled as class 1 due to the class one has smooth area, and because the speckle noise filter will remove important information. On the other hand, we have used the Wiener filter for the cases labelled class 1. That process will help to remove noise from any images that require filtering.

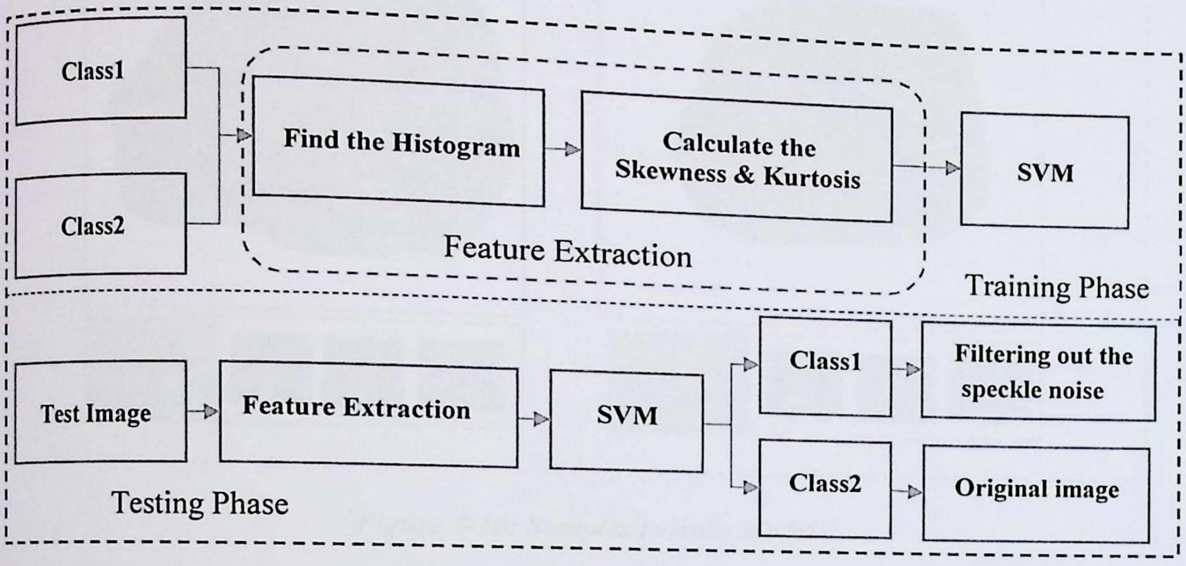



*Figure 7-9: Model 1: speckle noise based on the entire image*

### 7.3.2 Model 2: Adaptive speckle noise based on images blocks

In some cases, the images have two types of the region in the same image: solid and not solid textures (Figure 7-1-C). In this case, we have built an adaptive model to process the image based on blocks. Two phases are used to identify the region with a solid texture. Phase one is training, and phase two is testing. In the training step, the system takes 100 block samples from class one (solid texture) and 100 block samples from class two (not solid) as input, where the size of the block is $(30 \times 30)$ (see Figure 7-10). Then, the histogram for each individual block is found, followed by calculating the statics measurements, skewness and kurtosis, as based on the histogram determined in each instance. These statics measurements have been used to train a pattern classifier to identify the solid texture class from others (not solid class).

The testing phase aims to detect solid texture in out-of-sample images. The proposed model slides a fixed size blocks over the input ultrasound image and uses the trained SVM classifier to decide which patterns show solidity, or otherwise, in the blocks. for each window, the same set of features as in the training phase have been extracted to test the trained SVM classifier; the SVM predication determines whether the window needs to be processed.

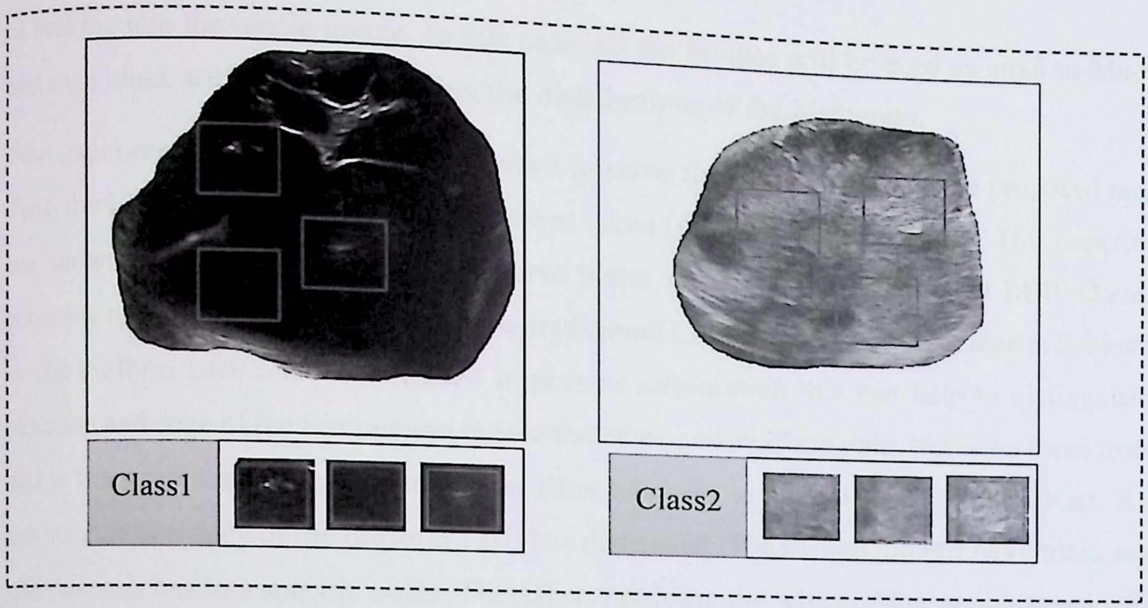



*Figure 7-10: Samples to train Model 2*

### 7.3.3 Experiments and Results

In order to evaluate the effects of the proposed models in this study, the Local Binary Pattern (LBP) 59 bins and LBP 255 bins have been used. These features are used to identify tumour type (benign or malignant) (Khazendar, et al. 2015). Therefore, we have used these features to evaluate our new approaches to pre-processing and enhancement, as extracted from the cropped ROI from the ovarian ultrasound images and fed to the neural network.

In this study, 242 two-dimensional images of ovarian tumours have been taken from a total of 232 patients. These images consisted of 104 malignant tumours and 138 benign tumours which were obtained by utilizing the two-dimensional B-mode feature. From each tumour class, 50 images were chosen randomly to give a total sample size of 100, which was particularly important given that benign tumours outnumbered malignant tumours (138 benign vs. 104 malignant). The *randsample* function in MATLAB was used for the unsystematic sampling. The parameters for ANN were set at 60% training (50% training, 10% validation) and 40% testing, and were then used to calculate sensitivity, accuracy and specificity. An average for each of these factors was found over 20 repetitions.

Figure 7-11 and Figure 7-12 show the effects of the two proposed models by using two texture features, namely LBP 59 bins and LBP 255 bins. We have tested both features with and without pre-processing. We note that pre-pressing based on local information has improved the accuracy of both features, notably in a model that works locally by looking at small regions instead of looking into the whole image. In this case, all the images will be used as input to Model 2 and each block will process regarding the distributions of the histogram.

Four experiments sets have been conducted to show the effectivenees of the proposed model. First, the LBP of the original image has been taken (without pre-processing). This experiment has shown that the uniform LBP (59) gave better result than the traditional LBP where the accuracy of uniform LBP is 80% and the traditional LBP is 74%. The main reason behind that is, the uniform LBP has captured the important information that can help to distinguish the structure and edge of the tumour and ignore the other non-uniform data. Based on these features and in the second experiment, the winner filter has been applied on the whole data set. As we can see, the accuracy of the uniform LBP has decreased. The benign tumour has a black region and this area has less speckle noise. Therefore, applying the winner filter on these regions will remove a lot of the important information's that can help identify the benign cases. This result has motived us to build an adaptive model that can help recognize the images with black region before we send it to the winner filter. In the third experiment, we have used the new proposed adoptive model to identify the cases with black area automatically from the cases that have rich textures. The winner filter applied on the image which has rich texture. This model has helped to identify the cases that have a big area of the reach textures and black area. The reason of reducing the accuracy of that model in some cases have different regions (rich textures and black region). Therefore, in the fourth experiment we have enhanced the proposed model to work with local information, i.e. divide the image into small blocks and see if the blocks need a filtering process or not. The result has shown that the adoptive model has enhanced the result for both features (uniform LBP and the LBP).

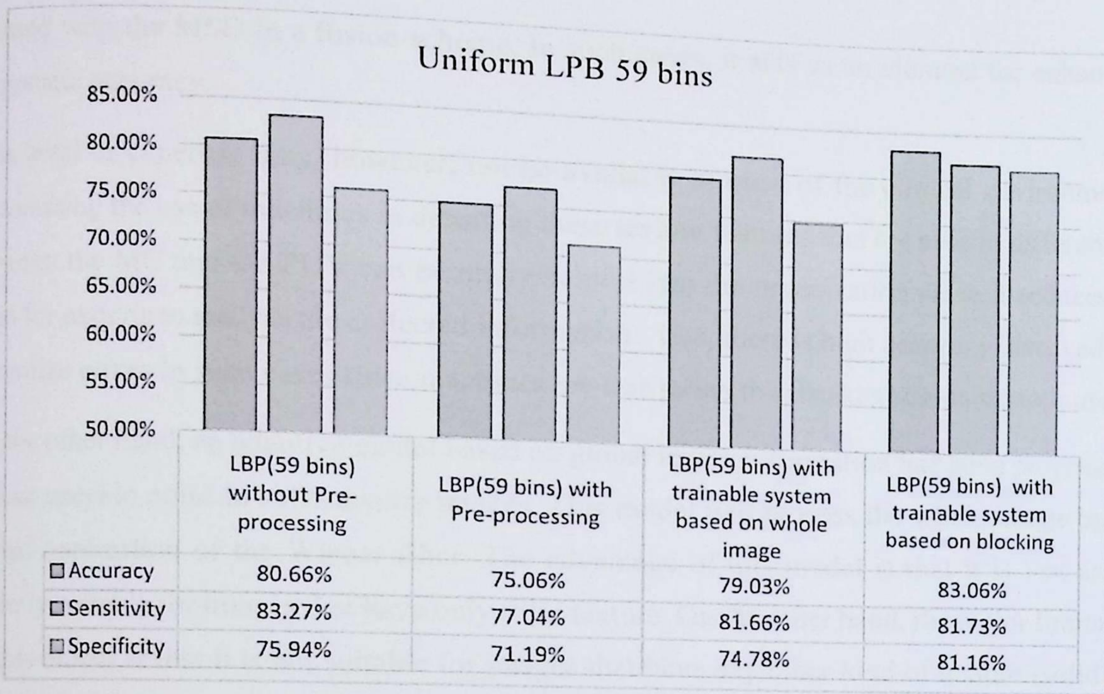



*Figure 7-11: Evaluation of the performance of Model 1 and Model 2 using uniform LBP (59 bins)*

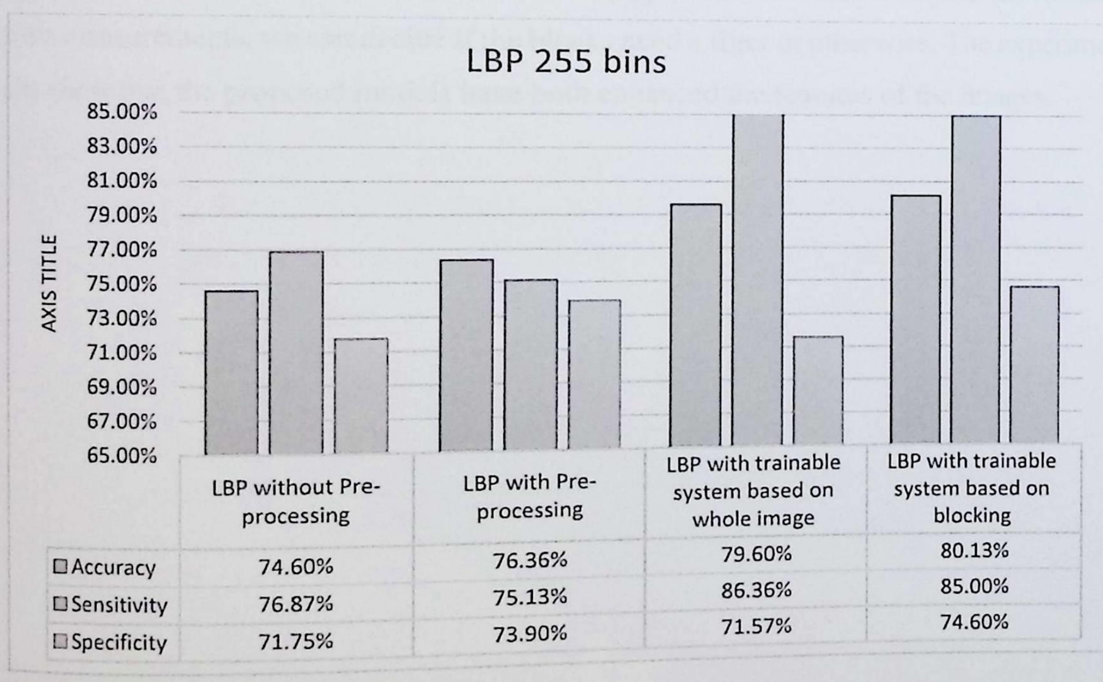



*Figure 7-12: Evaluation of the performance of Model 1 and Model 2 using LBP (255 bins)*

## 7.4 Summary

In summary, this chapter has shown that computer learning machines that use texture-based methodology have the capability to characterize static ultrasound images shown in cases of miscarriage. As shown in the studies, it can be concluded that the texture features have some extra value that has been added in order to discriminate between the MC and PUV. They can be used with the MSD in a fusion scheme. In such cases, it acts as an element for enhancing diagnostic accuracy.

This level of expertise may, however, not be available in most of the clinical environments. Automating the use of machines in detecting these texture features that are able to differentiate between the MC and the PUV can be really helpful into the organization since it reduces the need for experts to analyse the collected information. Also, there is high accuracy involved and minimize errors in such cases since machines are less prone to mistakes compared to humans. On the other hand, an adaptive model based on global image information has been proposed to reduce speckle noise in solid texture images. This model will process the whole image by the global application of the Wiener filter. The advantage of this model is that it is fast and is entirely suitable for images that have only solid texture. On the other hand, the major limitation of this model is that it is not suitable for images that have any other kind of texture (solid and not solid). Therefore, we have proposed a second model which is based on the use of local information. In this model, we will process the image as blocks. Based on the skewness and kurtosis measurements, we can decide if the blocks need a filter or otherwise. The experimental results show that the proposed models have both enhanced the features of the images.

# Chapter 8. Conclusion and Future work

The research work conducted for this thesis was designed to automate, using computer algorithms, the detection of gynaecological abnormalities from Ultrasound scans of the ovary. This is meant to exploit the rapid advances in computer technology to provide tools that assist clinician in making/confirming reliable decisions and thereby mitigating the effect of acute shortage of adequately skilled staff and in return reduce the growing cost of healthcare. The initial focus was on diagnosing miscarriage in early pregnancy but it was also implicit that any knowledge gained from that component of the work needs to be extended to related tasks in diagnosing ovarian tumour masses into benign and malignant. Identifying the main tasks in this project was fairly straightforward, but designing and developing computer algorithms to achieve the objectives of the individual tasks was anything but straightforward. The journey towards the development and evaluation of efficiently reliable solutions was, nevertheless, very fruitful and revealed many challenges that are specific to Ultrasound images but can arise in many other medical imaging modalities. This research project was conducted in collaboration with prominent UK and European medical experts as part of an effort to eventually develop a decision support system in this field, and their constant help and advice was instrumental in achieving the reported solutions.

This chapter highlights the main findings of the research conducted in each of the identified tasks and draw conclusions and implications for research beyond the stated objectives. Then, some of the recommendations, research directions and suggestions to be addressed in future work are outlined in order to further improve the performance and enhance the reliability of the proposed systems in analysing the ultrasound image images.

## 8.1 Conclusion

The use of ultrasound imagery to monitor abnormalities of the ovary is ubiquitous in modern clinics and is the key for the identification of potential problems in pregnancy and the ovarian tumour. Dimensional measurements of the Gestation sack (GS) and the Yoke Sack (YS) inside the GS including the MSD of the sac and the relative size of the YS are among the diagnostic indicators for miscarriage used by clinicions. In practice, MSD value in certain ranges is used as a criterion for diagnosing miscarriage while enlarged YS can also indicate pathology. In addition, there are different parameters (mentioned in chapter 2) that are used by oncologists to identify abnormalities in ovarian tumours. Therefore, the objective of the first component of this research was automatic segmentation of the GS (and YS) prior to computing these parameters. Image segmentation is a common task in image analysis, and a variety of automatic (or semi-automatic) solutions have been developed for different types of images. The success of such algorithms depends very much on how significant the variation in the distribution of image features (e.g. pixel intensity, waveform frequencies, entropy, and their filtered values), and expectedly their performances are influenced by the nature of the imaging modality and the objects of interest beside factors influencing image quality.

Existing segmentation methods i.e. threshold, region growing, watershed transform, and edge detection techniques have been shown to have limited use for the ultrasound images investigated tasks, due to the challenges associated with these tasks that were highlighted in each chapter (i.e. quality of the ultrasound, poor border of the ROI and the similarity between the ROI and the background). Testing the performance of these existing schemes, confirmed an important expected conclusion that there is no one method that could work well all the time for every imaging modality. Therefore, it became necessary to develop some pre- and post–segmentation techniques to avoid the limitations that we faced with existing methods.

Taking into account that speckle noise is widely believed to have an adverse influence on the performance of the existing methods, and the observed visual characteristics of wavelet transformed US images we developed an automatic segmentation that was based on thresholding post speckle de-noising in the wavelet domain. Wavelet transforms help visually highlight regions that have more similarity with the GS region and tumour masses, Despite the relatively high performance, that step resulted in many cases were the scheme detected few additional non-ROI object that exhibited similar properties. Moreover, our experiments revealed that speckle noise filters may remove the sought-after ROI (the GS) when it is relatively small size. To remove the additional Non-ROI detected objects we have developed an object detection model based on regularity geometry features. This procedure helped identify the GS object and remove the non-GS objects obtain in the previous step. Our evaluation of this version did not overcome all the limitations and some Non-ROI regions and the ROI have similar geometry/texture features. Therefore, in some cases, the proposed object detection model failed to identify the correct ROI an thereby resulting in some false positive cases.

The factors that led to the limitations of that version of our automatic segmentation scheme led to complementing this scheme with a machine learning component to help distinguish between

ROI and non-ROI for a more effective solution. This was motivated by the recent success stories on using the machine learning and the texture features in different applications. Therefore, we developed a trainable segmentation component that uses machine learning of the distinguishing texture features which works in 2 stages: 1) binaries the image to capture small objects as well as identify poor borders and 2) filter out Non-ROI objects depending on the distinguishing texture features.

Although this was very successful in identifying the sought after ROI, we found that in some cases it may not succeed always to capture the whole ROI (under-segmentation). This was finally overcome by adding a region growing as a post-segmentation procedure using texture features and the ANN (Trainable Region Growing) to extract the whole ROI. The trainable Region Growing method has reduce the under-segmentation problem.

Instead of a one-step general purpose segmentation, we found that the challenge of reliably segmenting the GS, and identifying the YS inside it, required a multi-level segmentation method that involve machine learning of certain texture features that at each stage the solution improves the segmentation by removing irrelevant objects and/or closing the border of the GS. Whereby at each to extract the ROI from the ovary ultrasound images. The completing of this task enabled the move to the next task of automatically measuring the MSD of the GS, and using the computed MSD to diagnose miscarriage in US pregnancy scans.

Having shown that the last version of the trainable segmentation successfully extracts ROIs (GS, and the YS inside it), it was necessary to demonstrate that our machine training has generated a model that is at least as good as the trained human clinician in relation to measuring the MSD. The closeness of the MSD measurements made by the clinicians (i.e. the manual ground truth) to the MSD values output by our automatic solution demonstrated a strong correlation between the two, and thereby confirmed that machine learning is capable of generating an automatic scheme that perform as well as trained human.

Due to the fact that analysis of Ovarian US scan images for diagnosing abnormalities exhibited by ovarian cysts, the issue of automatically segmenting these cysts is expected to benefit from the above trainable segmentation scheme of the GS. Unlike the case of pregnancy, where there is one GS object, in the case of tumour there may include more than one cyst. As observed in chapter 5, the presence of unilocular cyst and multilocular cysts is one of the influencing factors used by IOTA to distinguish between benign and malignant tumours. Furthermore, at penultimate stages of our automated GS trainable segmentation we get a number of ROIs. Hence, we applied and tested the above method to segment ovarian masses. The success of the various steps of the segmentation algorithm for this application confirmed that complementing of threshold based segmentation with machine learning approaches provides a perfect approach to overcome limitations caused by difficulty in determining features that can visibly distinguish ROI from non-ROI. Indeed, the use of ANN based region growing is testimony to the importance of this complementary approach.

Considering the initial objective of automating the computing the MSD measurements of the GS and the way machine learning influenced the developed complementary approach for accurate GS and ovarian masses segmentation raised an important question whether we could exploit further the use of machine learning for diagnostics without relying on ROI measurements. In fact, comparison of the decision made by the proposed method with the ground truth provided by the expert revealed that the MSD-based initial diagnosis agreement with ultimate diagnosis was high compared to the diagnosis when the computed MSD sits comfortably within the medically practiced criteria. This raised a new research question that motivated us to use machine learning as a new abnormality signature to complement the MSD-based decision. The new signature was based on training commonly used texture features within the segmented ROI to be used as a model for diagnostics. Again, our experimental work confirmed the success of this approach has demonstrated that the adopted texture feature can be used as a good indictor to identify the miscarriage in early stage. In fact, these results open the way to search for other digital image features that could further enhance accuracy and reduce both false positives and false negatives.

Finally, the way we developed the proposed hybrid multi-stage solutions, together with their highly successful outcomes presented in the various chapters do not only show the effectiveness of the proposed models but rather revealed the toughness of the challenges automatically analysing ultrasound images. But we have also established that machine learning provides an excellent set of tools to confront these and similar tasks in analysing biomedical images for diagnostics purposes.

## 8.2 Future Work

Achieving such a complex goal requires that the scope of our reported work be further extended to cover cases beyond those considered here. Ideally, future work should be dedicated to the improvement and further development of the current proposed methods to increase overall accuracy. In addition, the work should continue to extract important measurements/indicators from the ovarian ultrasound image (miscarriage and ovarian tumour). In particular, we will explore the following areas that still need additional work.

### 8.2.1 Miscarriage Ultrasound Segmentation

1- As mentioned in chapter 2, there are three stages related to pregnancy (Stage 1: empty GS, Stage 2: YS inside the GS and stage 3: Embryo inside the GS) (See Figure 8-1). We have only implemented the identification of the stages of pregnancy for both stage 1 and stage 2 due to the limited number of images available for stage 3. It would be both interesting and beneficial to investigate the method for all three stages of pregnancy. Secondly, the implementation of an effective model to segment, measure and classify of the embryo ultrasound image (stage 3) would also be beneficial, as this would help to identify miscarriage in the early stages of pregnancy.

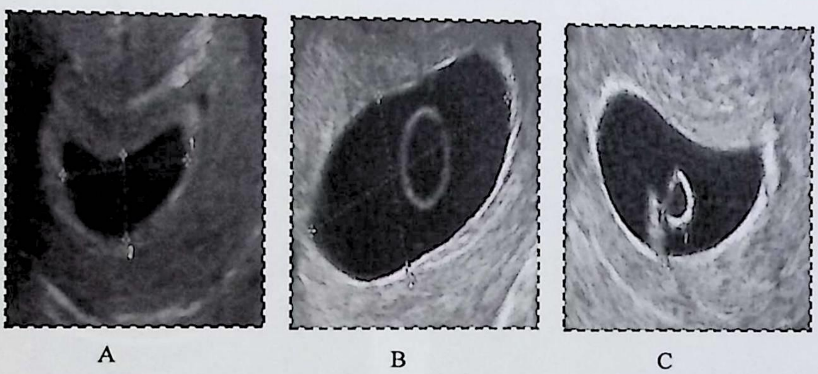



*Figure 8-1: Examples of ultrasound images of the very beginning of pregnancy until developing the embryo (A) gestational sac (B) YS within the GS (C) embryo attached with YS within the GS.*

2- Extend the machine learning approach beyond the cases tackled here. This requires capturing significantly more ultrasound scan images related to miscarriage cases to be used as considerable training and testing dataset necessary] for building reliably and scalable machine learning tools. Ideally, such a model should enhance TP rates as well as filter out non-ROIs.

### 8.2.2 Ovarian Tumour Segmentation

We are convinced of the potential of building on the success of using the machine learning based trainable GS segmentation, for the identification of identifying and distinguishing different types of ovarian tumours. In fact, in chapter 4 we proposed an effective algorithm to determine the nature of cyst locularity. Each type of tumour is characterised by: 1) cyst locations (**Unilocular, Multilocular**); 2) the presence/absence of acoustic shadows; 3) number of papillary projections; 4) solid tumour (**None, small, large**); and 5) internal wall structure (**Regular, Irregular, and Indeterminate**) (see Figure 8-2). It would be interesting to implement a model to extract these characteristics and use them as features with which to identify tumour types. In particular, we plan to investigate and develop tools, inspired by the achieved success, to differentiate between different types of benign, borderline and malignant cases.

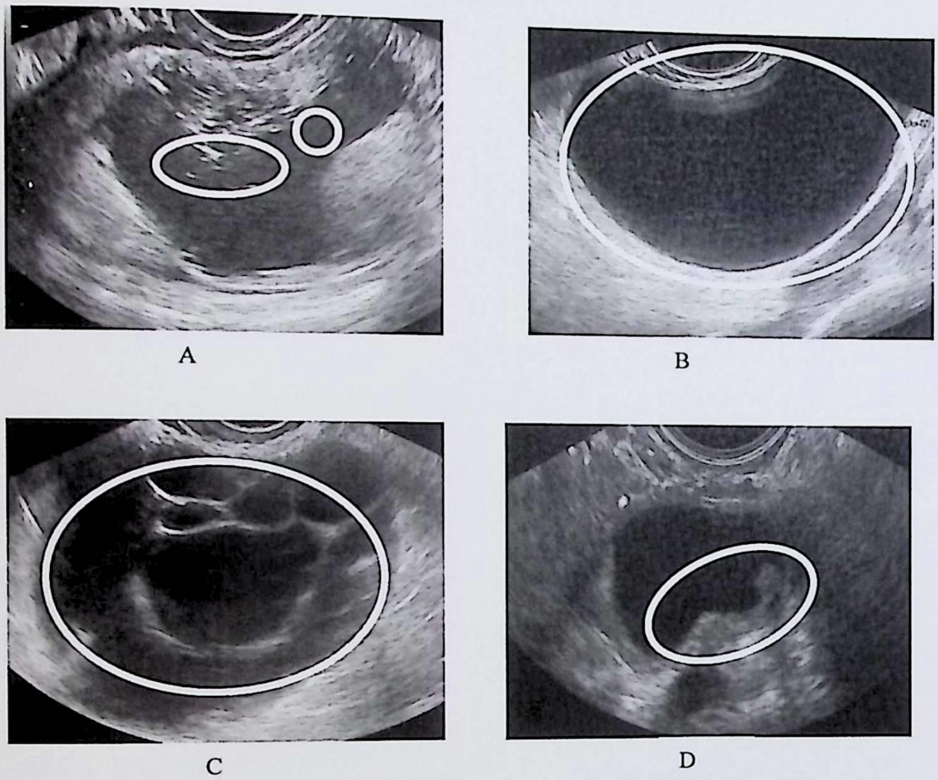



*Figure 8-2: Ovarian tumour signs: A) this type of tumour has clear signs, which are white dots and lines, B) this case includes acoustic shadows (present), unilocular, non-solid tumour and internal wall structure (regular), C) this includes multilocular signs and*

### 8.2.3 Ultrasound image Segmentation methods in general

1- Extend the applicability of the developed solutions to different diseases, i.e., prostate cancer and evaluate the effects of these models. Based on the succus of our trainable multi-level segmentation, we plan to tweak the proposed model to be suitable for the segmentation of the ROI in different applications.

2- Investigate further features and use the feature selection techniques to produce an effective trainable segmentation model that can reduce false positives to near zero.

3- The speckle noise and the quality of the ultrasound image represent the greatest challenge to the segmentation issue. Therefore, it would be interesting to build an effective model by using the spatial and frequency domains to reduce speckle noise in order to enhance image quality and highlight the ROI. This process should help to make the segmentation process easy and fast.

4- It would be interesting to investigate the use of three-dimensional ultrasound scans of the gestational sac to allow for more accurate segmentation of the GS and estimate volume and other parameters.

## References


Kapur, J. N., P. K. Sahoo, and A.K. C. Wong. 1985. "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram." *CVGIP* 29: 273-285.

Abdallah, Y, Anneleen Daemen, E Kirk, Anne Pexsters, O Naji, C Stalder, D Gould, et al. 2011. "Limitations of current definitions of miscarriage using mean gestational sac diameter and crown--rump length measurements: a multicenter observational study." *Ultrasound in Obstetrics \& Gynecology* (Wiley Online Library) 38 (5): 497-502.

Abdallah, Y, Anneleen Daemen, S Guha, S Syed, O Naji, Anne Pexsters, E Kirk, et al. 2011. "Gestational sac and embryonic growth are not useful as criteria to define miscarriage: a multicenter observational study." *Ultrasound in Obstetrics \& Gynecology* (Wiley Online Library) 38 (5): 503-509.

Abrahim, Banazier A, Zeinab A Mustafa, Inas A Yassine, Nourhan Zayed, and Yasser M Kadah. 2012. "Hybrid total variation and wavelet thresholding speckle reduction for medical ultrasound imaging." *Journal of Medical Imaging and Health Informatics* (American Scientific Publishers) 2 (2): 114-124.

Albregtsen, Fritz, and others. 2008. "Statistical texture measures computed from gray level coocurrence matrices." *Image processing laboratory, department of informatics, university of oslo* 5.

Al-Fahdawi, Shumoos, Rami Qahwaji, Alaa S Al-Waisy, Stanley Ipson, Rayaz A Malik, Arun Brahma, and Xin Chen. 2016. "A fully automatic nerve segmentation and morphometric parameter quantification system for early diagnosis of diabetic neuropathy in corneal images." *computer methods and programs in biomedicine* (Elsevier) 135: 151-166.

AL-Hassan, Nadia. 2014. "Face recognition in uncontrolled conition-super resolution and compressive sensing." Ph.D. dissertation, The University of Buckingham.

Alrawi, AzmiTawfik, Ali Makki Sagheer, and Dheyaa Ahmed Ibrahim. 2012. "Texture segmentation based on multifractal dimension." *International Journal on Soft Computing* (Academy \& Industry Research Collaboration Center (AIRCC)) 3 (1): 1.

Ambedkar, B. 2012. *Ultrasonic Coal-Wash for De-Ashing and De-Sulfurization: Experimental Investigation and Mechanistic Modeling*. Springer Science \& Business Media.

American cancer society. 2014. 5 August. Accessed September 3, 2015. http://www.cancer.org.

Anqing, Yang. 2010. "Research on image filtering method to combine mathematics morphology with adaptive median filter." (IET).

Aramendía-Vidaurreta V, Cabeza R, Villanueva A Navallas J Alcázar JL Juan Luis. 2016. "Ultrasound Image Discrimination between Benign and Malignant Adnexal Masses Based on a Neural Network Approach." *Ultrasound in medicine \& biology* (Elsevier) 42 (3): 742-752.

Bezdek, James C, LO Hall, and LP Clarke. 1993. "Review of MR image segmentation techniques using pattern recognition." *Medical physics* (Wiley Online Library) 20 (4): 1033-1048.

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

Bourne, Tom. 2016. "Why greater emphasis must be given to getting the diagnosis right: the example of miscarriage." *Australasian Journal of Ultrasound in Medicine* (Wiley Online Library) 19 (1): 3-5.

Braem, Marieke GM, N Charlotte Onland-Moret, Leo J Schouten, Roy FPM Kruitwagen, Annekatrin Lukanova, Naomi E Allen, Petra A Wark, et al. 2012. "Multiple miscarriages are associated with the risk of ovarian cancer: results from the European Prospective Investigation into Cancer and Nutrition." *PloS one* (Public Library of Science) 7 (5): e37141.

Buades, Antoni, Bartomeu Coll, and Jean Michel Morel. 2004. "On image denoising methods." *CMLA Preprint* 5.

Bushberg, Jerrold T, J Anthony Seibert, Edwin M Leidholdt Jr, and John M Boone. 2002. "The Essential Physics of Medical Imaging, Lippincott Williams \& Wilkins." *Philadelphia, USA.*

2015. *Cancer Research UK* . 17 june. Accessed Septermber 2015. http://www.cancerresearchuk.org.

Cancer stats key facts . 2011. 1 Septermber . Accessed Septermber 5, 2015. www.cancerreserach.uk.org.

Carneiro, Gustavo, Bogdan Georgescu, Sara Good, and Dorin Comaniciu. 2008. "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree." *IEEE transactions on medical imaging* (IEEE) 27 (9): 1342-1355.

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

Chen, Yan, Shira L Broschat, and Patrick J Flynn. 1996. "Phase insensitive homomorphic image processing for speckle reduction." *Ultrasonic imaging* (SAGE Publications Sage CA: Los Angeles, CA) 18 (2): 122-139.

Clarke, LAURENCE P. 1996. "Partially supervised clustering for image segmentation." *Pattern recognition* (Citeseer) 29 (5): 859-871.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-vector networks." *Machine learning* (Springer) 20 (3): 273-297.

Costa, Alceu Ferraz, Gabriel Humpire-Mamani, and Agma Juci Machado Traina. 2012. "An efficient algorithm for fractal analysis of textures." *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images.* 39-46.

Dahdouh, Sonia, Elsa D Angelini, Gilles Grange, and Isabelle Bloch. 2015. "Segmentation of embryonic and fetal 3D ultrasound images based on pixel intensity distributions and shape priors." *Medical image analysis* (Elsevier) 24 (1): 255-268.

Dalal, Navneet, and Bill Triggs. 2005. "Histograms of oriented gradients for human detection." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* 886-893.

David H. Hareva, Irene A. Lazarusli, and Suryasari. 2016. "Automatic Gestational Age Estimation Based on Crown Rump Length and Gestational Sac." *Journal of Image and Graphics* 4.

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

Dirk Timmerman, Lil Valentin and Tom Bourne. 1999. *International Ovarian Tumor Analysis 2018.* Accessed 2018. http://www.iotagroup.org/.

Doubilet, Peter M, Carol B Benson, Tom Bourne, and Michael Blaivas. 2013. "Diagnostic criteria for nonviable pregnancy early in the first trimester." *New England Journal of Medicine* (Mass Medical Soc) 369 (15): 1443-1451.

2012. *Ectopic pregnancy and miscarriage: diagnosis and initial management.* National Institute for Health and Care Excellence (NICE) of Ectopic Pregnancy and. Accessed 2015.

Egmont-Petersen, Michael, Dick de Ridder, and Heinz Handels. 2002. "Image processing with neural networks—a review." *Pattern recognition* (Elsevier) 35 (10): 2279-2301.

Fasihi, Maedeh Sadat, and Wasfy B Mikhael. 2016. "Overview of Current Biomedical Image Segmentation Methods." *Computational Science and Computational Intelligence (CSCI), 2016 International Conference on.* 803-808.

Gai, Shan, Boyu Zhang, Cihui Yang, and Lei Yu. 2018. "Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution." *Digital Signal Processing* (Elsevier) 72: 192-207.

Gangeh, Mehrdad J, Hadi Tadayyon, Lakshmanan Sannachi, Ali Sadeghi-Naini, William T Tran, and Gregory J Czarnota. 2016. "Computer aided theragnosis using quantitative ultrasound spectroscopy and maximum mean discrepancy in locally advanced breast cancer." *IEEE transactions on medical imaging* (IEEE) 35 (3): 778-790.

Garg, Amit, and Vineet Khandelwal. 2018. "Combination of Spatial Domain Filters for Speckle Noise Reduction in Ultrasound Medical Images." *Advances in Electrical and Electronic Engineering* 15 (5): 857-865.

Geirsson, RT, and RMC Busby-Earle. 1991. "Certain dates may not provide a reliable estimate of gestational age." *BJOG: An International Journal of Obstetrics \& Gynaecology* (Wiley Online Library) 98 (1): 108-109.

Giakoumelou, Sevi, Nick Wheelhouse, Kate Cuschieri, Gary Entrican, Sarah EM Howie, and Andrew W Horne. 2015. "The role of infection in miscarriage." *Human reproduction update* (Oxford University Press) 22 (1): 116-133.

Gonzalez, Rafael C., and Richard E. Woods. 2017. *Digital Image Processing.* 4th. Pearson International Edition.

Grau, Vicente, AUJ Mewes, M Alcaniz, Ron Kikinis, and Simon K Warfield. 2004. "Improved watershed transform for medical image segmentation using prior information." *IEEE transactions on medical imaging* (IEEE) 23 (4): 447-458.

Gustavo, Bogdan Georgescu, Sara Good, and Dorin Comaniciu. 2007. "Automatic fetal measurements in ultrasound using constrained probabilistic boosting tree." *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 571-579.

Hamid, Bidi Ab. 2011. "Image texture analysis of transvaginal ultrasound in monitoring ovarian cancer." Ph.D. dissertation, Cardiff University.

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

Hwang, Kyung-Hoon, Jun Gu Lee, Jong Hyo Kim, Hyung-Ji Lee, Kyong-Sik Om, Minki Yoon, and Wonsick Choe. 2005. "Computer aided diagnosis (CAD) of breast mass on ultrasonography and scintimammography." *Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005. Proceedings of 7th International Workshop on.* 187-189.

Imaduddin, Zaki, Muhammad Ali Akbar, Hilmy Abidzar Tawakal, I Putu Satwika, and Yudianto B Saroyo. 2015. "Automatic detection and measurement of fetal biometrics to determine the gestational age." *Information and Communication Technology (ICoICT), 2015 3rd International Conference on.* 608-612.

Jung, Chanho, and Changick Kim. 2010. "Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization." *IEEE transactions on biomedical engineering* (IEEE) 57 (10): 2600-2604.

Karthikeyan, K, and C Chandrasekar. 2011. "Speckle noise reduction of medical ultrasound images using Bayesshrink wavelet threshold." *International Journal of Computer Applications* (International Journal of Computer Applications, 244 5 th Avenue,\# 1526, New York, NY 10001, USA India) 22 (9): 8-14.

Kasban, H, MAM El-Bendary, and DH Salama. 2015. "A Comparative Study of Medical Imaging Techniques." *International Journal of Information Science and Intelligent System.*

Kaur, Amandeep, and Karamjeet Singh. 2010. "Speckle noise reduction by using wavelets." *National Conference on Computational Instrumentation CSIO NCCI.* 198-203.

Khazendar, S, A Sayasneh, H Al-Assam, Helen Du, Jeroen Kaijser, L Ferrara, Dirk Timmerman, S Jassim, and Tom Bourne. 2015. "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 (1): 7.

Khazendar, Shan, Hisham Al-Assam, Tom Bourne, and Sabah A Jassim. 2014. "Automatic Identification of Early Miscarriage Based on Multiple Features Extracted From Ultrasound Images."

Khazendar, Shan, Jessica Farren, Hisham Al-Assam, Hongbo Du, Ahmed Sayasneh, Tom Bourne, and Sabah Jassim. 2015. "Automatic Identification of Miscarriage Cases Supported by Decision Strength Using Ultrasound Images of the Gestational Sac."

Krivanek, Anthony, and Milan Sonka. 1998. "Ovarian ultrasound image analysis: Follicle segmentation." *IEEE transactions on medical imaging* (IEEE) 17 (6): 935-944.

Kumar, Dhrub, Maitreyee Dutta, and Parveen Lehana. 2013. "A Comparative Analysis of Different Wavelets for Enhancing Medical Ultrasound Images." *International Journal of Computer Applications* (Citeseer) 66 (7).

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

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

Levi, CS, EA Lyons, and DJ Lindsay. 1990. "Ultrasound in the first trimester of pregnancy." *Radiologic clinics of North America* 28 (1): 19-38.

Loizou, Christos P, Constantinos S Pattichis, Christodoulos Christodoulou, Robert SH Istepanian, Marios Pantziaris, Andrew Nicolaides, and others. 2005. "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery."

*Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on* (IEEE) 52 (10): 1653-1669.

Loizou, Christos P,. 2005. "Ultrasound image analysis of the carotid artery." Ph.D. dissertation, Imperial College.

Loizou, P, and Constantinos S Pattichis. 2008. "Despeckle filtering algorithms and software for ultrasound imaging." *Synthesis lectures on algorithms and software in engineering* (Morgan \& Claypool Publishers) 1 (1): 1-166.

Loupas, T, WN McDicken, and PL Allan. 1989. "An adaptive weighted median filter for speckle suppression in medical ultrasonic images." *Circuits and Systems, IEEE Transactions on* (IEEE) 36 (1): 129-135.

Lu, Wei, Jinglu Tan, and Randall Floyd. 2005. "Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform." *Ultrasound in medicine and biology* (Elsevier) 31 (7): 929-936.

Malathi, G, and V Shanthi. 2010. "Histogram based classification of ultrasound images of placenta." *IJCA*. 975-8887.

Martin, David, Charless Fowlkes, Doron Tal, and Jitendra Malik. 2001. "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics." *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on.* 416-423.

Mateo, Juan L, and Antonio Fernandez-Caballero. 2009. "Finding out general tendencies in speckle noise reduction in ultrasound images." *Expert systems with applications* (Elsevier) 36 (4): 7786-7797.

McInerney, Tim, Ghassan Hamarneh, Martha Shenton, and Demetri Terzopoulos. 2002. "Deformable organisms for automatic medical image analysis." *Medical image analysis* (Elsevier) 6 (3): 251-266.

Meiburger, Kristen M, U Rajendra Acharya, and Filippo Molinari. 2017. "Automated localization and segmentation techniques for B-mode ultrasound images: A review." *Computers in biology and medicine* (Elsevier).

Meyer, Fernand. 1994. "Topographic distance and watershed lines." *Signal processing* (Elsevier) 38 (1): 113-125.

Michailovich, Oleg V, and Allen Tannenbaum. 2006. "Despeckling of medical ultrasound images." *ieee transactions on ultrasonics, ferroelectrics, and frequency control* (IEEE) 53 (1): 64-78.

Moghaddam, Mostafa Jabarouti, and Hamid Soltanian-Zadeh. 2011. "Medical image segmentation using artificial neural networks." In *Artificial Neural Networks-Methodological Advances and Biomedical Applications*. InTech.

Mohanaiah, P, P Sathyanarayana, and L GuruKumar. 2013. "Image texture feature extraction using GLCM approach." *International Journal of Scientific and Research Publications* (Citeseer) 3 (5): 1.

Morris, DT. 1988. "An evaluation of the use of texture measurements for the tissue characterisation of ultrasonic images of in vivo human placentae." *Ultrasound in medicine \& biology* (Elsevier) 14 (5): 387-395.

Nebraska, Omaha. 2013. *Diagnostic and Interventional Radiology Department / Ultrasound of Early Pregnancy.* Creighton University Medical Center. http://web.archive.org/web/20070814054851/http:/radiology.creighton.edu/pregnancy.htm#section4.

NHS, N. H. S. 2015. *A miscarriage is the loss of a pregnancy during the first 23 weeks.* http://www.nhs.uk/Conditions/Miscarriage/Pages/Introduction.aspx.

Ni, Bo, Fa-zhi He, Yi-teng Pan, and Zhi-yong Yuan. 2016. "Using shapes correlation for active contour segmentation of uterine fibroid ultrasound images in computer-aided therapy." *Applied Mathematics-A Journal of Chinese Universities* (Springer) 31 (1): 37-52.

Ni, Bo, Fazhi, and ZhiYong Yuan. 2015. "Segmentation of uterine fibroid ultrasound images using a dynamic statistical shape model in HIFU therapy." *Computerized Medical Imaging and Graphics* (Elsevier) 46: 302-314.

Ni, Dong, Yong Yang, Shengli Li, Jing Qin, Shuyuan Ouyang, Tianfu Wang, and Pheng Ann Heng. 2013. "Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters." *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on.* 772-775.

Nixon, Mark, and Aguado Alberto. 2012. *Feature Extraction and Image processing.* 3nd. Elsevier.

Noble, J Alison, and Djamal Boukerroui. 2006. "Ultrasound image segmentation: a survey." *IEEE Transactions on medical imaging* (IEEE) 25 (8): 987-1010.

Ojala, Timo, Matti Pietikäinen, and David Harwood. 1996. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition* (Elsevier) 29 (1): 51-59.

Ojalaa, Timo, Matti Pietikainen, and Topi Maenpaa. 2002. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Transactions on pattern analysis and machine intelligence* (IEEE) 24 (7): 971-987.

Otoum, Nesreen. 2013. "Medical image processing: applications in ophthalmology and total hip replacement." Ph.D. dissertation, {\copyright} Nesreen Otoum.

Otsu, N. 1975. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man and Cybernetics* 9 (1): 62-66.

Otsu, Nobuyuki. 1979. "A threshold selection method from gray-level histograms." *EEE transactions on systems, man, and cybernetics* 9: 62-66.

Padmapriya, B, and T Kesavamurthy. 2016. "Detection of follicles in poly cystic ovarian syndrome in ultrasound images using morphological operations." *Journal of Medical Imaging and Health Informatics* (American Scientific Publishers) 6 (1): 240-243.

Patidar, Pawan, Manoj Gupta, Sumit Srivastava, and Ashok Kumar Nagawat. 2010. "Image de-noising by various filters for different noise." *International journal of computer applications* (International Journal of Computer Applications, 244 5 th Avenue,\# 1526, New York, NY 10001, USA India) 9 (4).

Patil, Mrs Smita S, Mrs AA Junnarkar, and Ms DV Gore. 2014. "Study of Texture Representation Techniques." *image* 3 (3).

Pham, Dzung L, Chenyang Xu, and Jerry L Prince. 2000. "Current methods in medical image segmentation." *Annual review of biomedical engineering* (Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA) 2 (1): 315-337.

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

Potocnik, Bozidar, and Damjan Zazula. 2000. "Automated ovarian follicle segmentation using region growing." *Image and Signal Processing and Analysis, 2000. IWISPA 2000. Proceedings of the First International Workshop on.* 157-162.

Preisler, Jessica, Julia Kopeika, Laure Ismail, Veluppillai Vathanan, Jessica Farren, Yazan Abdallah, Parijat Battacharjee, et al. 2015. "Defining safe criteria to diagnose miscarriage: prospective observational multicentre study." (British Medical Journal Publishing Group).

Rabbani, Hossein, Mansur Vafadust, Purang Abolmaesumi, and Saeed Gazor. 2008. "Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors." *Biomedical Engineering, IEEE Transactions on* (IEEE) 55 (9): 2152-2160.

Rangayyan, Rangaraj M. 2005. *Biomedical Image Analysis.* 1st. CRC Press.

Rawat, Vidhi, Alok Jain, Vibhakar Shrimali, and Abhishek Rawat. 2013. "Automatic assessment of foetal biometric parameter using GVF snakes." *International Journal of Biomedical Engineering and Technology* (Inderscience Publishers Ltd) 12 (4): 321-333.

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

Saranya, R, and S Uma Maheswari. 2016. "Follicle detection in ovary image using adaptive particle swarm optimization." *Journal of Medical Imaging and Health Informatics* (American Scientific Publishers) 6 (1): 125-132.

Sayasneh, A, L Ferrara, Bavo De Cock, S Saso, M Al-Memar, S Johnson, J Kaijser, et al. 2016. "Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study." *British journal of cancer* (Nature Publishing Group) 115 (5): 542.

Sayasneh, Ahmad, Christine Ekechi, Laura Ferrara, Jeroen Kaijser, Catriona Stalder, Shyamaly Sur, Dirk Timmerman, and Tom Bourne. 2015. "The characteristic ultrasound features of specific types of ovarian pathology." *International journal of oncology* (Spandidos Publications) 46 (2): 445-458.

Schapire, Robert E. 2013. "Explaining adaboost." In *Empirical inference*, 37-52. Springer.

Shan, Juan. 2011. *A fully automatic segmentation method for breast ultrasound images.* Utah State University.

Sharma, Mona, and Sameer Singh. 2001. "Evaluation of texture methods for image analysis." *Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001.* 117-121.

Sharma, Neeraj, and Lalit M Aggarwal. 2010. "Automated medical image segmentation techniques." *Journal of medical physics/Association of Medical Physicists of India* (Medknow Publications) 35 (1): 3.

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

Soille, P. 2013. *Morphological image analysis: principles and applications.* Springer Science & Business Media.

Stolojescu-crisan, Cristina, and Ştefan Holban. 2013. "A Comparison of X-Ray Image Segmentation Techniques." *Advances in Electrical and Computer Engineering* 13: 85-92.

Straka, Matus, Alexandra La Cruz, A Kochl, Milos Sramek, E Groller, and Dominik Fleischmann. 2003. "3D watershed transform combined with a probabilistic atlas for medical image segmentation." *Journal of Medical Informatics \& Technologies* 6: IT69--78.

Sudha, S, GR Suresh, and R Sukanesh. 2009. "Speckle noise reduction in ultrasound images by wavelet thresholding based on weighted variance." *International Journal of Computer Theory and Engineering* (IACSIT Press) 1 (1): 7-12.

Supriyanti, Retno, Ahmad Abdul Hafidh, Yogi Ramadhani, and Haris B Widodo. 2018. "MEASURING GESTATIONAL AGE AND UTERINE DIAMETER BASED ON IMAGE SEGMENTATION." *ARPN Journal of Engineering and Applied Sciences.*

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

Thaipanich, Tanaphol, and C-C Jay Kuo. 2010. "An adaptive nonlocal means scheme for medical image denoising." *Medical Imaging 2010: Image Processing.* 76230M.

The Minitab Blog. 2016. *Regression Analysis.* Accessed December 31, 2016. http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit.

Tomar, Divya, and Sonali Agarwal. 2015. "Twin support vector machine: a review from 2007 to 2014." *Egyptian Informatics Journal* (Elsevier) 16 (1): 55-69.

Torloni, Maria Regina, N Vedmedovska, M Merialdi, AP Betran, T Allen, R Gonzalez, and LD Platt. 2009. "Safety of ultrasonography in pregnancy: WHO systematic review of the literature and meta-analysis." *Ultrasound in Obstetrics and Gynecology* (Wiley Online Library) 33 (5): 599-608.

UIDELINE, LALPRATI. 2005. "Ultrasound evaluation of first trimester pregnancy complications." *J Obstet Gynaecol Can* 27 (6): 581-585.

V C, Hough Paul. 1962. "Method and means for recognizing complex patterns." *U.S. Patent 3,069,654.*

Van Calster, Ben, Kirsten Van Hoorde, Lil Valentin, Antonia C Testa, Daniela Fischerova, Caroline Van Holsbeke, Luca Savelli, et al. 2014. "Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study." *Bmj* (British Medical Journal Publishing Group) 349: g5920.

Viola, Paul, and Michael Jones. 2001. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.* I--I.

Wells, William M, W Eric L Grimson, Ron Kikinis, and Ferenc A Jolesz. 1996. "Adaptive segmentation of MRI data." *IEEE transactions on medical imaging* (IEEE) 15 (4): 429-442.

WiseGeek. 2013. *Medical Imaging.* http://www.wisegeekhealth.com/what-is-medical-imaging.htm.

Wu, Shibin, Qingsong Zhu, and Yaoqin Xie. 2013. "Evaluation of various speckle reduction filters on medical ultrasound images." *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE.* 1148-1151.

Yu, Jinhua, Yuanyuan Wang, Ping Chen, and Yuzhong Shen. 2008. "Fetal abdominal contour extraction and measurement in ultrasound images." *Ultrasound in medicine and biology* (Elsevier) 34 (2): 169-182.

Zhang, Dong, Yu Liu, Yan Yang, Menglong Xu, Yu Yan, and Qianqing Qin. 2016. "A region-based segmentation method for ultrasound images in HIFU therapy." *Medical physics* (Wiley Online Library) 43 (6Part1): 2975-2989.

Zhang, Ling, Siping Chen, Chien Ting Chin, Tianfu Wang, and Shengli Li. 2012. "Intelligent scanning: Automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination." *Medical physics* (Wiley Online Library) 39 (8): 5015-5027.

Zhanga, Ling, Siping Chen, Shengli Li, and Tianfu Wangb. 2011. "Automatic measurement of early gestational sac diameters from one scan session." *Proc. of SPIE Vol.* 796342-1.

Zheng, Quanquan, Yingjie Liu, and Weiliang Zhu. 2013. "Uterine calcifications segmentation and extraction from ultrasound images based on level set." *Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on.* 591-594.

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

Zumray Dokur, Tamer Olmez. 2002. "Segmentation of ultrasound images by using a hybrid neural network." *Pattern Recognition Letters* (Elsevier) 23 (14): 1825-1836.