Annals of the BMVA Vol. 2015, No. 5, pp 1−16 (2015)

Automatic Identification of Miscarriage Cases Supported by Decision Strength Using Ultrasound Images of the Gestational Sac

Shan Khazendar[1]
Jessica Farren[2], Hisham Al-Assam[1], Hongbo Du[1], Ahmed Sayasneh[2], Tom Bourne[2], Sabah Jassim[1]
[1] Department of Applied Computing, University of Buckingham, Buckingham, UK.
[2] Department of Early Pregnancy, Imperial College, Queen Charlotte's and Chelsea Hospital, London, UK.
<shan.khazendar@buckingham.ac.uk>, < jessicafarren@doctors.org.uk >

Abstract

Ultrasound imaging is one of the most widely used multipurpose imaging modalities for monitoring and diagnosing early pregnancy events. The first sign and measurable element of an early pregnancy is the appearance of the Gestational Sac (GS). Currently, the size of the GS is manually estimated from ultrasound images. The manual measurements tend to result in inter- and intra-observer variations, which may lead to difficulties in diagnosis. This paper proposes a new method for automatic identification of miscarriage cases in the first trimester of pregnancy. The proposed method automatically segments the GS and calculates the Mean Sac Diameter (MSD) and other geometric features of the segmented sac. After classifying the image based on the extracted features into either a pregnancy of unknown viability (PUV) or a possible miscarriage case, we assign the decision with a strength level to reflect its reliability. The paper argues that the level of decision strength gives more insight into decision making than other classical alternatives and makes the automated decision process closer to the diagnosis practice by experts.

1 Introduction

Medical imaging techniques have been increasingly deployed in the past few decades to assist diagnoses of various types of diseases. Among medical imaging modalities, ultrasound imaging is considered to be safe, non-invasive, portable, accurate, and cost effective. These advantages have made ultrasound imaging the most common diagnosis tool deployed in hospitals around the world [Michailovich and Tannenbaum, (2006)]. Ultrasound imaging is also considered as an effective modality particularly for monitoring

pregnancy because of its safety without radiation hazard [Geiresson and Busby-Earle, (1991)].

The first three months, known as the first trimester, are the most crucial period in pregnancy [Joseph et al. (2012)]. Monitoring within this period enables clinics to evaluate the development, growth, and wellbeing of the foetus [Kaur and Kaur, (2011)]. The first sign and measurable element of early pregnancy is the GS. The American College of Radiology guideline defines miscarriage as being an empty GS with a MSD greater than or equal to 16 mm [Levis et al. (1990)]. A recent study reviewed this cut-off [NICE clinical guideline, (2012)] and concluded that an empty GS with MSD greater than or equal to 25 mm should be introduced as a new guideline in order to minimize the risk of false positive diagnosis of miscarriage [NICE clinical guideline, (2012)]. This is irrespective of the assumed gestation of the pregnancy (calculated from the last menstrual period). If the empty gestation sac does not meet these diagnostic criteria for a miscarriage, a repeat scan is suggested in 7-10 days. In the interim, it is described as a pregnancy of unknown viability (PUV) – denoting the fact that there is neither evidence of viability, nor conclusive evidence of miscarriage.

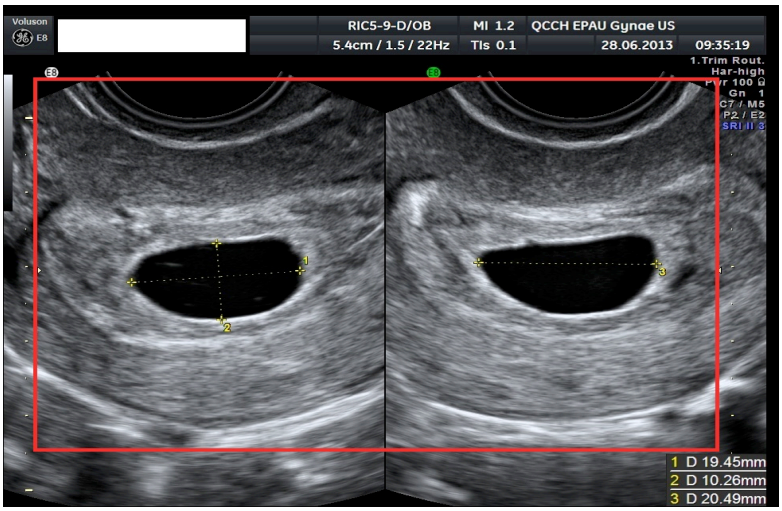
Estimating the size of the GS is currently done manually. The manual process involves multiple subjective decisions when the three diameter measurements on the GS, explained in section 3.3, are taken from static ultrasound images in 2 planes, to calculate the MSD [NICE clinical guideline, (2012)]. The subjective decisions increase the inter- and intra-observer variations [Pexsters et al. (2011)] which may lead to errors at diagnosis, and requires there to be margin of safety in the diagnostic criteria chosen. An automated way of estimating the size of the GS from a given ultrasound image is therefore desirable. Automatic measurement of the size of the GS requires segmenting the GS from a given ultrasound image first. Unlike other types of medical images, ultrasound images are corrupted by speckle noise that tends to reduce image quality and contrast, and consequently increases the level of difficulty in segmenting GS. A considerable amount of research into ultrasound image de-nosing has been undertaken [Jawad, (2007), Hiremath et al. (2009), Hiremath and Tegnoor, (2010)], but the research in automatic GS segmentation is limited. In [Chakkarwar et al. (2010)], a method that uses a combination of contrast enhancement, low pass filtering and Wiener filtering to de-speckle the image, followed by thresholding, was reported with an average accuracy of 83.3% over a small database of 12 images. In [Zhang et al. (2011)] an algorithm for detecting GS from a video was proposed based on using the AdaBoost method to identify the GS from each frame followed by exploiting the local context and eliminating false positive detections. The algorithm was tested on 31 videos and achieved a GS detection rate of 87.5%.

In [Khazendar et al. (2014)], we reported a method to automatically segment and estimate the size of the GS in terms of the MSD. We used the k Nearest Neighbour (kNN) [Du, (2010)] classifier to identify early miscarriage cases based on the automatically measured MSD. We also compared the classification performance of using MSD with other geometric features from GS images such as volume, perimeter, area, circularity, compactness, solidity and eccentricity on a limited dataset of 68 images. This paper extends our work presented in [Khazendar et al. (2014)] in two ways. First, we expand the existing dataset with more recently acquired images totalling 184 images to evaluate our

method and consolidate our findings. Second, we introduce the concept of decision strength levels into the classification stage. We argue that the level of decision strength provides more insight than other classical classification methods and makes the proposed decision making closer to the practical diagnosis of miscarriage cases by experts. Our experimental results using the 184 images confirm that the proposed solution achieves a high level of accuracy using the automatically estimated MSD as well as the perimeter and volume of the sac.

2 Miscarriage Diagnosis Based on Manual Measurements

An ultrasound machine is a real-time imaging device which can scan the region of interest using different probe angles. Gynaecologists usually scan the image in the sagittal plane, i.e. the vertical longitudinal plane of the uterus, locate the GS, and select two points on the boundaries of the GS using joy sticks to measure its diameters (major and minor). Afterwards, they change the probe angle by 90 degrees to capture the image in the transverse plane, i.e. the horizontal plane that is perpendicular to the coronal and sagittal planes, and then takes the third diameter measurement. The MSD is defined as the average of these three measurements. Diagnosis is made according to the refined NICE guideline [NICE clinical guideline, (2012)]: if the MSD is greater than or equal to 25 mm with no obvious contents (Yolk Sac (YS) or embryo) inside, miscarriage is pronounced. Fig 1 presents an example image of the GS taken in respective sagittal and transverse planes. The red rectangle represents the main fan area. The GS is the dark region in the center. The area outside the red rectangle is called margin area. It shows information about the patient (blocked for anonymity), the date and time that the image was taken and the ultrasound machine setting. The figure shows the three manual measurements of the GS size marked by yellow dotted lines. The measurement results in millimeters are present in the margin area. There are other signs of likely miscarriage. For example, the border of the GS appears irregular in its shape. Although a GS growth abnormality is within our scope of investigation, this paper is only concerned with the miscarriage diagnosis based on the GS size measurements.



Sagittal plane view Transverse plane view

Figure 1: Ultrasound image of GS in Sagittal and Transverse planes

3 The Proposed Method for Automatic Miscarriage Diagnosis

3.1 Materials

The ultrasound images used in our study were obtained in two batches. The first dataset contains 94 ultrasound images: 79 images are PUV cases and 15 images of miscarriage cases. The second independently sampled dataset contains 90 images among which 78 images are PUV cases and 12 are of miscarriage cases. All images are taken at various points of time in the first trimester of pregnancy, collected and labelled by the author (JF) in the Early Pregnancy Units, Imperial College Healthcare Trust, London, UK. Each image consists of two views of a GS from two perpendicular sections/planes as explained in section 2.

3.2 Methods

Fig 2 shows the block diagram of the underlying process of the proposed solution for automatic identification of miscarriage cases. The process consists of a sequence of stages starting from cropping a region of interest, followed by enhancing the image, segmenting GS from the enhanced image, extracting diameter measurements of best fit ellipse shape, and finally classifying the GS as miscarriage or PUV based on the [NICE clinical guideline, (2012)], with a classification strength indicator. Each stage of the process will be explained in details in the following sub-sections.

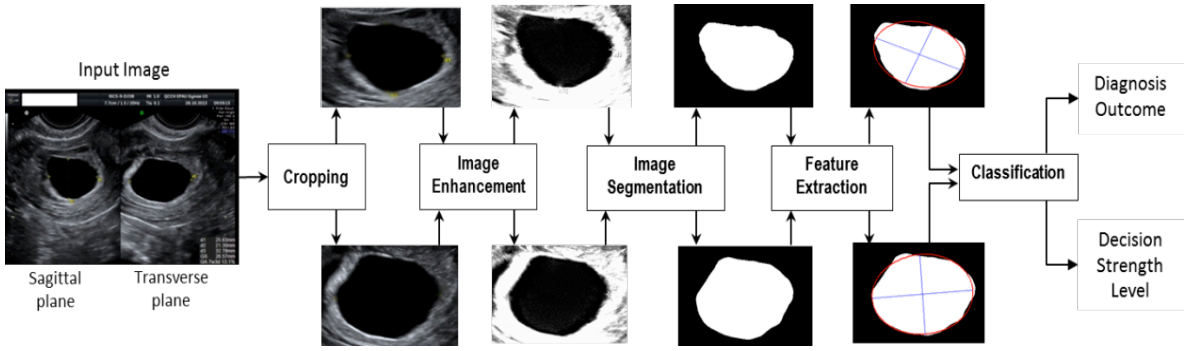


Figure 2: Block Diagram of the Major Steps of the Proposed Method

3.2.1 Image Cropping

Each original input image contains two-plane views of the GS with the margin areas. Before the GS segmentation starts, the margin areas of both views need to be removed by using the imcrop function in Matlab with a fixed position vector parameter (30, 150, 900, 500) where the first two components represent the co-ordinates of the top-left corner point and the latter two components represent the width and height of the crop region. The cropped image is as shown in Fig 3(a). Then we separate the two views from the middle of the image. The resulting two images are shown in Fig 3(b).

3.2.2 Image Enhancement

Ultrasound images of the GS are typically dark, causing difficulties in segmenting the GS. We used the following heuristics filter to enhance the image:

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

where $I(i,j)$ is the intensity value of the pixel at (i, j) position, and μ is the mean of all pixel intensity values. Unlike histogram equalisation, this simple pixel value transformation gives more weight to dark pixels where the detailed information lays by stretching them over the whole grey-scale range. The main aim of this enhancement is to highlight the GS for ease of segmentation. The highlighted area of the GS is as shown in Fig 3(c).

3.2.3 GS Segmentation

The GS segmentation stage involves a series of operational steps due to the noisy surroundings near the sac. These steps are described as follows:

Step 1: Thresholding. The Otsu thresholding method [Otsu, (1979)] is first applied to the enhanced image to obtain a binary image as shown in Fig 3(d). In the binary image, the GSs as well as a number of false regions and small irrelevant objects are isolated from the background.

Step 2: Smoothing. A median filter with a window size of 15x15 is applied to the resulting image of step 1. As illustrated in Fig 3(e), this operation smooth the boundary of the sac without losing its original shape, fills small holes/gaps in the GS region, and helps connecting the non-sac or false regions to image borders for later removal.

Step 3: False regions removal. The imclearborder function in Matlab is then applied to clear all false regions that are connected to the image border, resulting in a clean image as shown in Fig 3(f).

Step 4: Further noise removal. Any small objects remaining in the image are considered as noises and should be removed. This is done by labelling each object using the Matlab function bwlabel, calculating the area of each object, and then deleting all small objects. The only remaining object is the GS as shown in Fig 3(g).

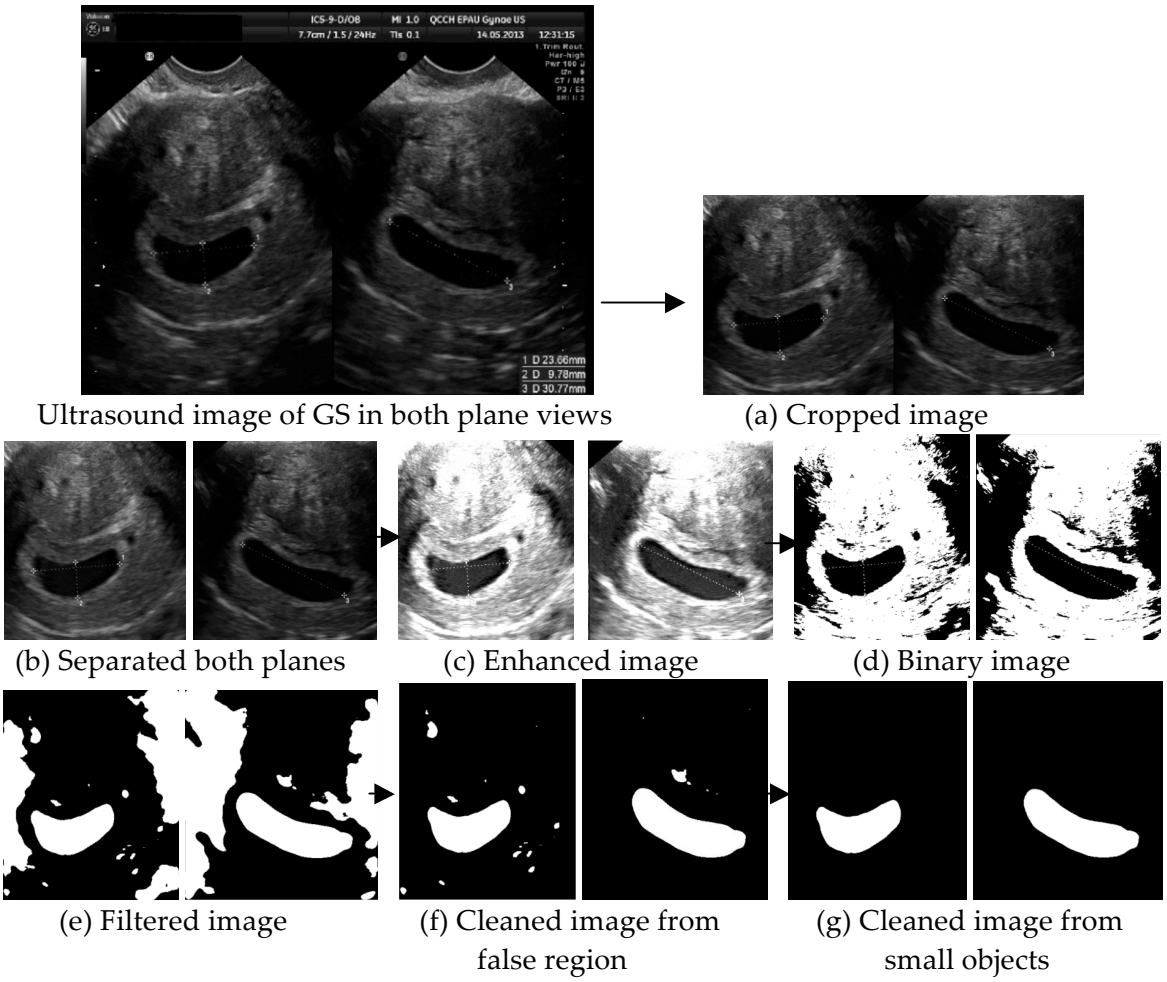


Figure 3: Steps of GS segmentation

3.2.4 Feature Extraction

As explained earlier, each GS is viewed in two perpendicular planes. The GS is usually round in shape at the early pregnancy, but as the sac grows it appears more elliptical. Therefore, our algorithm first finds the best fitting ellipse for the segmented GS in each plane. The regionprops function in Matlab is used to fit an ellipse to the sac region by matching the normalized second central moments based on [Haralick and Shapiro, (1992)]. This function returns four parameters: Major and Minor axes, Centroid and Orientation which is the angle between major and minor axes, as shown in Fig 4. Assuming the GS has an ellipsoidal shape in 3D, the three principal axes of the ellipsoid can be estimated by the major axis (A), minor axis (B) of the ellipse from the sagittal plane and the major axis (C) from the transverse plane.

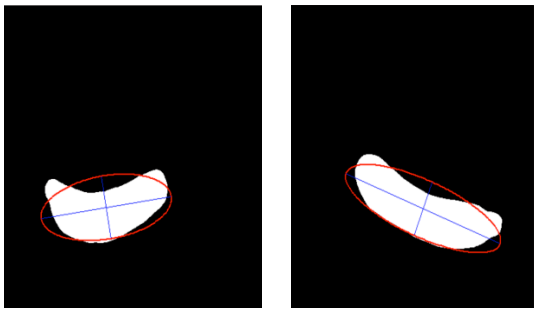


Figure 4: Best fitting ellipse for feature extraction

After that, we extract the following geometric features from each GS:

1. The Mean Sac Diameter (MSD): This is simply the average of three principal diameters along the A, B, and C axes.

2. Volume: The volume of the GS can be estimated using the three principal diameters as follows:
$$V = \frac{1}{8}.\frac{4}{3} \cdot \pi \cdot \mathrm{A} \cdot B \cdot C$$
We then take the cubic root of the volume as a volume measure.

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

For each image, the extracted features can be treated as separate features or as components of a feature vector for diagnosis purposes. To highlight the usefulness of these features at this point of the paper, we randomly selected 15 images of miscarriage cases and 15 images of PUV cases, and plot the automated measurements upon these features in Fig 5 The scatter plots show a clear separation of miscarriage and PUV cases.

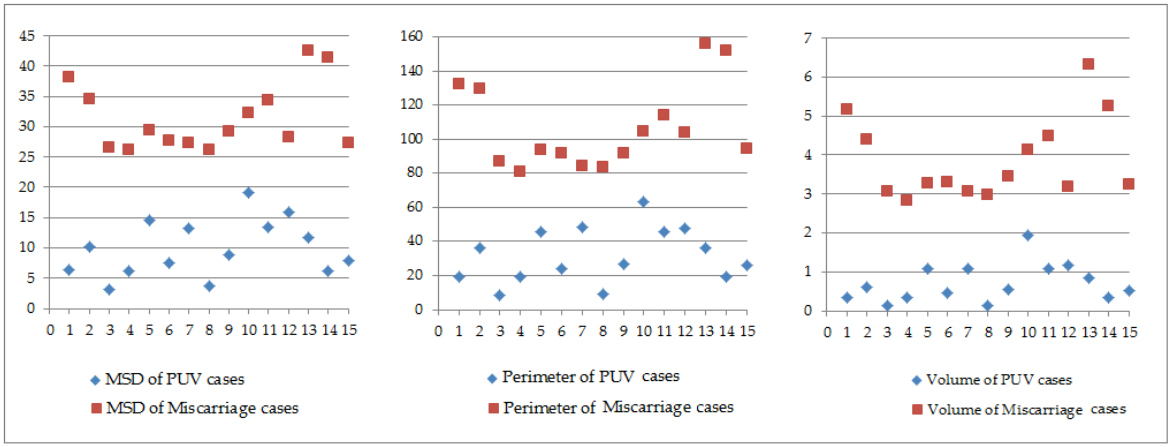


Figure 5: Scatter plots of automated measurements upon the three features for sample images

3.2.5 Classification

In principle, any appropriate classifier can be trained and deployed in this step. In this particular study, we used a simple kNN classifier to evaluate the effectiveness of our segmentation and feature extraction methods. The kNN classifier determines the class of a testing image by calculating the distances between the testing image's feature or feature vector and that of each exemplar image in the training set, locating the k nearest exemplars and using majority voting to decide the class of the testing image.

3.2.6 Level of Decision Strength

In our previous work [Khazendar et al. (2014)], we evaluated classification accuracy in terms of sensitivity and specificity. However, medical diagnosis in practice is not only concerned with whether a case is correctly diagnosed, but also how *strong* and *reliable* the diagnosis decision tends to be. We therefore propose a new and simple scheme to quantify the strength of a classification decision. Our proposed scheme goes through the following steps when classifying a testing image:

Step 1: Calculate the distance between the testing image and each exemplar image in the training set.

Step 2: Normalize the distance measurements from step 1 using the division-by-range method, then sort these distances in ascending order.

Step 3: Assign the class sign (- for miscarriage cases and + for PUV cases) to the distance measurements for the nearest k distances.

Step 4: Calculate the average of the three nearest signed distance measurements and use the average as a decision score.

Step 5: Assuming the distances from step 1 follow a normal distribution (around 65% of the distances falls within one standard deviation σ away from the mean μ), we determine the level of decision strength of classifying the testing image by using the decision score (Sc) from step 4 and the standard deviation (σ) of the distances' distribution according to the following three rules:

Rule 1: If $Sc(x) \geq (\sigma)$ or $Sc(x) \leq (-\sigma)$, then the decision strength is defined as **High**;
Rule 2: If $Sc(x) \geq (\sigma/2)$ and $Sc(x) < (\sigma)$) or ($Sc(x) \leq (-\sigma/2)$ and $(Sc(x) > (-\sigma))$, then the decision strength is defined as **Medium**;
Rule 3: If $Sc(x) < (\sigma/2)$ or $Sc(x) > (-\sigma/2)$, then the decision strength is deemed as **Low**.

The rationale behind this decision strength scheme is that the mean distance defines a near zero borderline between positive and negative classes. A distance-based decision is stronger if the distance is further away from the borderline. The thresholds we adopted in the scheme are heuristically determined.

4 Experiment Results & Discussion

4.1 Experimental Protocol

In this paper we conducted two sets of experiments to evaluate our automated miscarriage identification solutions. In the first set of experiments, we adopted an external test option widely adopted in clinical studies and trials: we used the first dataset (DS1) containing 94 images (79 PUV and 15 miscarriages) for training and the second dataset (DS2) of 90 images (78 PUV and 12 miscarriages) for testing (see section 3.1 for the acquisition of DS1 and DS2). To overcome the class imbalance problem between the PUV and miscarriage cases in the training set, we employed a down-sampling strategy as follows. In each round, we randomly selected 15 images from 79 images of PUV cases and combined them with the 15 images of miscarriage cases to form a training sample of 30 images. We repeated the test 15 times and then reported the average of accuracy, sensitivity and specificity.

In the second set of experiments, we adopted a typical machine learning test option by combining DS1 and DS2 into a single dataset of 184 images (157 PUV cases and 27 miscarriage cases). A stratified cross-validation was employed as the test protocol for this experiment. We also used a similar down-sampling strategy to resolve the class imbalance problem. Namely, a random sample was drawn 15 times. For each sample, 25 randomly selected (out of 157) PUV images and 25 randomly selected (out of 27) miscarriage images were chosen. We reported the average of accuracy, sensitivity and specificity.

We used the k Nearest Neighbor (kNN) classifier with k = 1 and k = 3 respectively for classification in both experiments. Besides, both individual features and the feature vector containing all three features as components were attempted. For each individual feature, the distance between two examples was measured as the absolute difference between their feature measurements. For the feature vector, the Euclidean distance was used as the distance function on the feature vectors of two examples.

4.2 Experimental Results

4.2.1 Automatic MSD vs. Manual MSD

To illustrate the effectiveness of our automatic segmentation, we compared the manual measurements of major and minor axes provided by domain experts with the automatic ones. Fig 6 presents a scatter plot of all 184 images as points along the automatically estimated and the manually measured MSD dimensions. The figure shows that the values of automatic MSDs are very close to manual measurements.

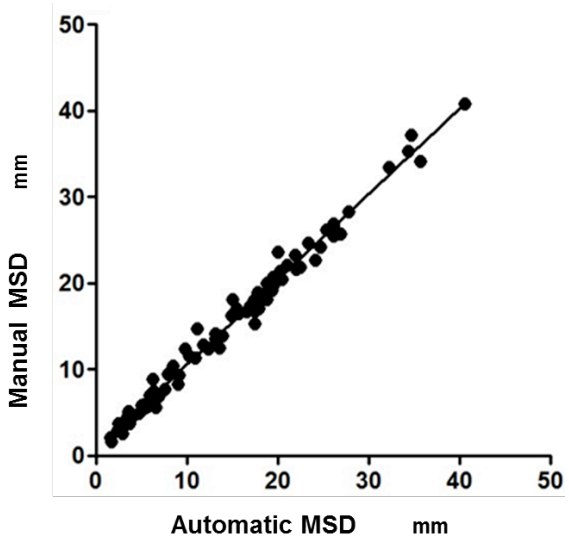
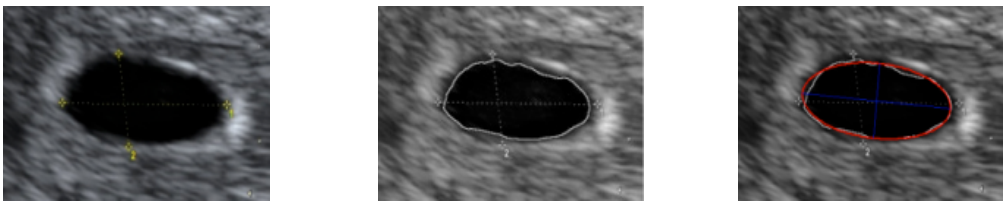


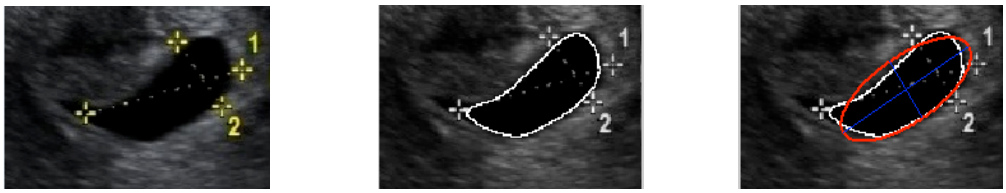
Figure 6: Differences between manual and automatic measurements of MSD (R- square= 0.98)

Fig 7 presents the images of the two example GSs with manual landmarks as well as automatic measurements of the two diameters in the sagittal plane (d1 and d2). The figure shows that both types of measurement lead to margins of error from the actual size of the GS. The automatic method always attempts to search for the maxima when estimating the diameters of the best-fit ellipsoid, whereas human operators do not and can be very subjective.



Manually measured d1 & d2 Detected Border of GS Estimated d1, d2 vs Manual d1, d2

(a) A regular shaped GS where manual and automatic measurements are similar



Manually measured d1 & d2 Detected Border of GS Estimated d1, d2 vs Manual d1, d2

(b) An irregular shaped GS where manual and automatic measurements are different

Figure 7: Examples of manual vs. automatic measurements

From further investigation of the PUV images that were misclassified, i.e. false positives, we found that the MSDs of these images, from 20 to 25 mm, are very close to the borderline between the PUV and miscarriage cases. The second author, an expert in clinical miscarriage diagnosis, confirmed that these borderline PUV cases all eventually became cases of miscarriage in follow-up diagnoses. Table 1 shows the borderline examples that were diagnosed by an expert and the automatic method. This suggests that our automatic method enabled more accurate and earlier prediction than the manual method in this case. More experiments are needed to consolidate this finding.

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

Table 1. Represents the Manual vs. Automatic diagnosis

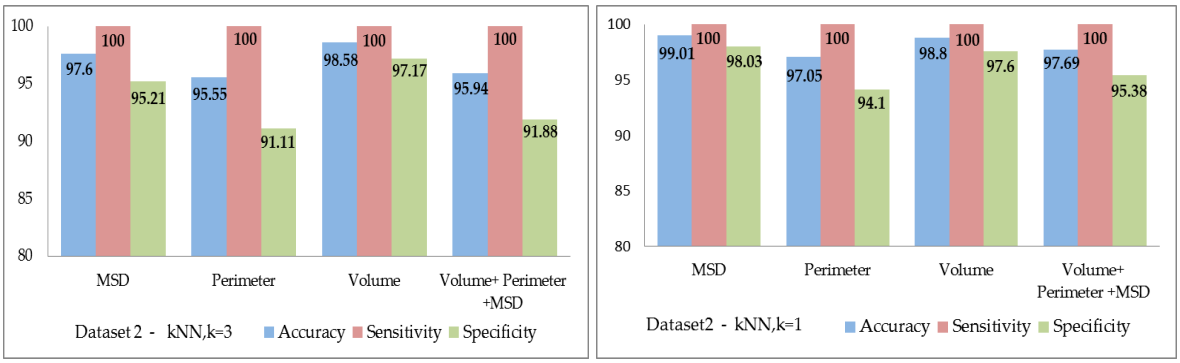
4.2.2 Classification Results & Discussions

This subsection presents the results of the two sets of experiments for evaluating the proposed method as explained in section 4.1. Fig 8(a) shows the result of the first set of experiments. When k = 3, using the automatic MSD feature alone, an overall accuracy near to 98% with sensitivity (miscarriage) of 100% and specificity (PUV) of 95% was obtained. This compares with an overall accuracy of 98% (sensitivity 100%, and specificity 97%) by using the volume feature alone, and an overall accuracy of 96% (sensitivity of 100% and specificity of 91%) by using the perimeter feature alone. Combining all three features does not seem to lead to better classification results. This is because the three types of features are highly correlated.

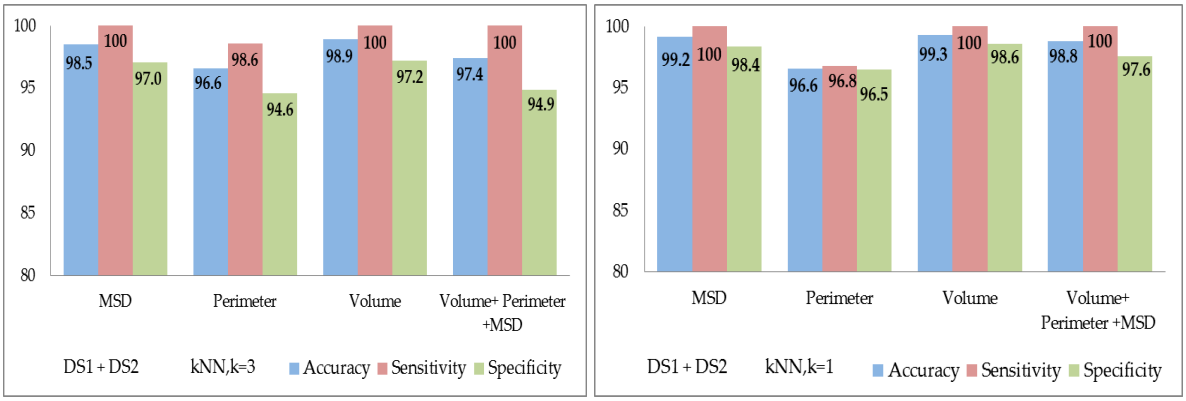
The results from the second set of experiments on the effectiveness of the extracted features give a similar reading as those from the first set of experiments. While the sensitivity remains high for MSD and volume features, the sensitivity for the perimeter feature is reduced marginally. At the same time, the specificity for each type of feature has

improved. Again, using the three types of features together in classification does not improve performance.

It should be said that the perimeter remains as a robust feature which is automatically measureable from each single image of the GS and still gives relatively good performance, and hence should not be easily dismissed. It should also be noted that the results of the two experiments project a slightly contradicting picture with the findings of our previous work reported in [Khazendar et al. (2014)], but we recognise that our previous findings were based on a rather small dataset (DS1).



(a) Results of the first set of experiments

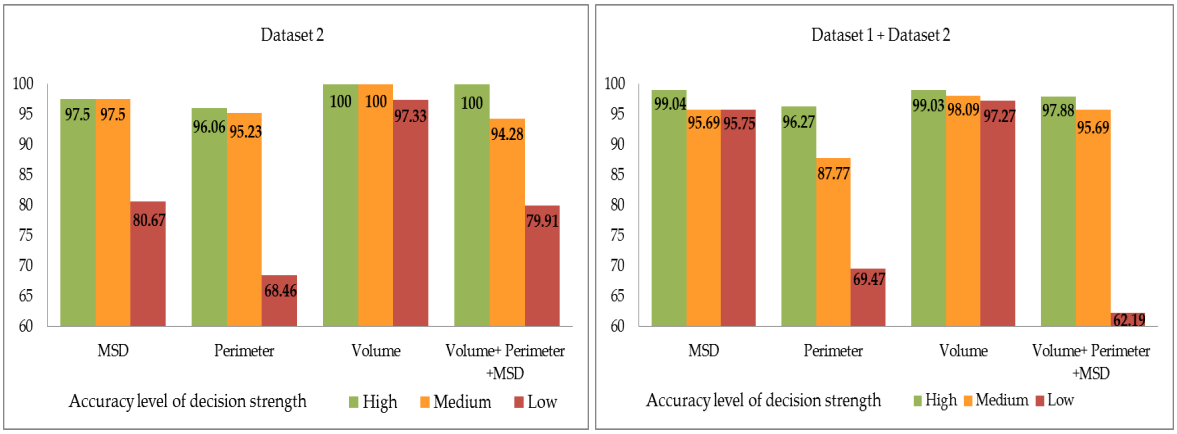


(b) Results of the second set of experiments

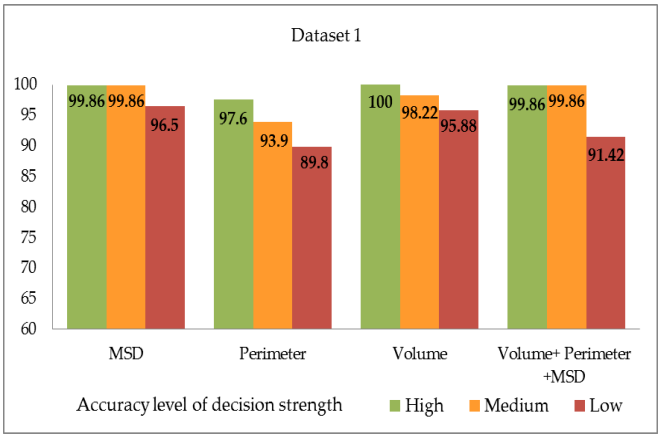
Figure 8: Comparison of miscarriage classification accuracy, sensitivity and specificity based on MSD, perimeter, volume and multi features using kNN, k=1,3

Fig 9 presents the results from the first set (Fig 9(a)) and second set of experiments (Fig 9(b)) when a classification decision is made with an associated level of strength. For comparison purposes, we also applied the strength-based classification scheme to DS1 the result shown in Fig 9(c). With the MSD feature in particular, the high decision strength leads to 97.5% overall accuracy, the medium decision strength also leads to the same level of accuracy whereas the decisions with low strength have a much lower level of accuracy of 80.6%. For the perimeter feature, the accuracy with low strength level can be as low as 68%. The results show that there is a correspondence between the decision accuracy and

the level of decision strength. In terms of coverage, when the MSD feature is used, for instance, among the total of 90 images in DS2, 39 images were classified with High level of strength, 40 images with Medium level strength, and 11 images with Low level strength.



(a) (b)



(c)

Figure 9: Accuracies of classifying images using different features at different levels of decision strength

After analyzing the classified images, we noticed that most of the images that have been classified with Low level of decision strength are those images where MSDs are measured between 13 mm and 26 mm, which are borderline cases (MSD between 16 mm and 25 mm) according to the guideline in [NICE clinical guideline, (2012)], and hence more challenging to diagnose in practice. This means that our decision strength levels correspond to the guideline's findings. Furthermore, most images that have been classified with High level of decision strength are of miscarriage cases whereas most images that have been classified with Low or Medium decision strength are PUV cases. This result might be desirable, but more work is needed to look into the factors that affect the outcomes.

We have also investigated other types of features such as area, compactness, circularity, solidity and eccentricity [Olver et al. (2010)]. Area is another indicator of GS size, and therefore it should have a performance close to that of the features described before. Compactness, solidity, circularity and eccentricity are indicators of roughness, irregularity, and the shape of GS borders. Fig 10 presents some initial findings with those features. The results suggest that comparing to MSD, perimeter and volume, these features alone have limited diagnostic value for miscarriage cases since most accuracies are close to 50%. One possible reason for the poor performance may be due to the fact that the images in our datasets have relatively regular and smooth borders. This means that we should include more images of various border characteristics in the datasets for our future studies. More investigations are needed to find the potential values for these features when they are collectively used in a fusion solution for miscarriage diagnosis.

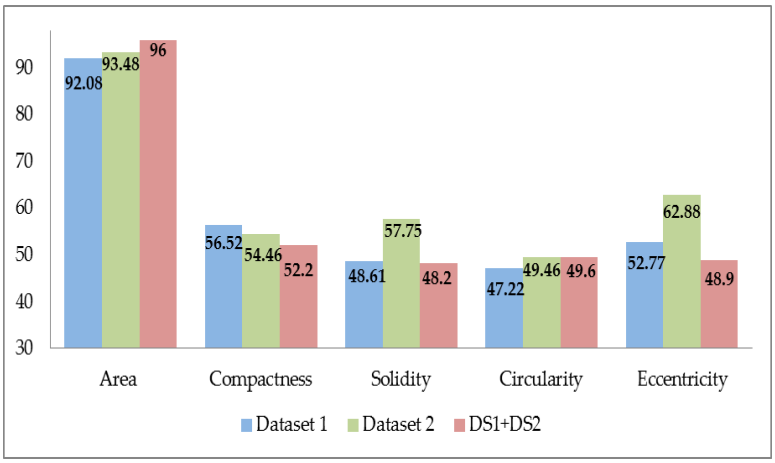


Figure 10: The system performance based on area, compactness, circularity, solidity and eccentricity

5 Conclusions

In this paper, an automatic computer-based solution for segmenting, quantifying and classifying empty GSs has been proposed for miscarriage diagnosis. First, our pre-processing operations employed in the solution are successful in segmenting the GS from noisy surroundings. Second, the MSDs automatically estimated by the solution are close to those of manual measurement, but more objective and deterministic than the subjective measurement. Third, we have investigated the effectiveness of new features such as perimeter and volume. According to our results we conclude that these two features lead to similar or better accuracy than MSD. Moreover, in this paper we have proposed a novel decision strength based classification scheme. The results have shown a correlation between the level of accuracy and the level of strength based decision. Such strength-based prediction outcomes provide more insight than only the predicted class. In particular, our finding between predictions with Low level of strength and borderline cases indicates that the outcome of our method is close to the diagnosis practice.

Our future work includes further investigations into other features, particularly features capturing the border characteristics of GS, and effects of external descriptors such as age and the level of human Chorionic Gonadotropin (hCG) in the framework of decision strength based classification of miscarriage cases using ultrasound images of GS.

Acknowledgements

The author SK wishes to thank the Ministry of the higher Education in *Kurdistan* for funding her PhD study.

References

O.V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Trans Ultrason Ferroelectr Freq Control*, 53(1):64-78, 2006.

R.T. Geirsson and R.M. Busby-Earle. Certain Dates may not provide a reliable estimate of gestational age. *BJOG: An International Journal of Obstetrics & Gynaecology*, 98(1):108-109, 1991.

H.K. Joseph, M.W. Guile, J.L. Bienstock, H.E. Fox, and E.E. Wallach. The Johns Hopkins manual of gynecology and obstetrics. *Lippincott Williams & Wilkins*, 2012.

A. Kaur and A. Kaur. Transvaginal ultrasonography in first trimester of pregnancy and its comparison with transabdominal ultrasonography. *Journal of Pharmacy and Bioallied Sciences*, 3(3):329-338, 2011.

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

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

National Institute for Health and Care Excellence (NICE) of Ectopic Pregnancy and Miscarriage. *Diagnosis and initial management in early pregnancy of ectopic pregnancy and miscarriage,* Guideline No. (154), 2012.

J. Awad. Prostate segmentation and regions of interest detection in transrectal ultrasound images. *PhD Thesis. University of Waterloo*, Canada, 2007.

P.S. Hiremath, P.T. Akkasaliga, and S. Badiger. Despeckling medical ultrasound images using the contourlet transform. *Indian International Conference on Artificial Intelligence – IICAI*, 1814-1827, 2009.

P.S. Hiremath and J.R. Tegnoor. Automated detection of follicle in ultrasound images of ovaries using edge based method. *Recent trends in image processing and pattern recognition (RTIPPR)*, (10):120-125, 2010.

A. Chakkarwar, S. Joshi, and S. Revankar. Automated analysis of gestational sac in medical image processing. *IEEE 2nd International Advance Computing Conference (IACC),* 304-309, 2010.

L. Zhang, S. Chen, S. Li and T. Wang. Automatic measurement of early gestational sac diameters from one scan session. *SPIE Medical Imaging*, *7963*, 2011.

N. Otsu. A threshold selection method from gray-level histograms. IEEE Transactions *on* Systems. Man *and* Cybernetics, (9):62-66, 1979.

H. Du. Data Mining Techniques and Applications: An Introduction, Course Technology Cengage Learning, 2010.

F. Olver, D. Boisvert and R. Clark. NIST handbook of mathematical functions. *Cambridge University Press*, 2010.

S. Khazendar, J. Farren, H. Al-assam, H. Du, A. Sayasneh, T. Bourne and S. Jassim. Automatic identification of early miscarriage based on multiple features extracted from ultrasound images. *Medical Image Understanding and Analysis (MIUA),* 131:136, 2014.

A. Pexsters, J. Luts, D. Van Schoubroeck, C. Bottomley, B. Van Calster, S. Van Huffel, Y. Abdallah, T. D'Hooghe, C. Lee, D. Timmerman and T. Bourne. Clinical implications of intra-and interobserver reproducibility of transvaginal sonographic measurement of gestational sac and crown–rump length at 6–9 weeks' gestation. *Ultrasound in obstetrics & gynecology*, 38(5): 510-515, 2011.

M. Haralick and L. Shapiro. Computer and robot vision. *Addison-Wesley*, 1992.