Automated Analysis of the Placenta in Ultrasound

by

Ricky Hu

B.A.Sc., The University of British Columbia, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL
STUDIES
(Biomedical Engineering)

The University of British Columbia
(Vancouver)

July 2019

© Ricky Hu, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

**Automated Analysis of the Placenta in Ultrasound**

submitted by **Ricky Hu** in partial fulfillment of the requirements for the degree of **Master of Applied Science in Biomedical Engineering**.

**Examinig Committee:**

Robert Rohling, Electrical and Computer Engineering  
*Supervisor*

Roger Tam, Biomedical Engineering  
*Supervisory Committee Member*

Shuo Tang, Electrical and Computer Engineering  
*Supervisory Committee Member*
Abstract

The placenta is an organ that serves as an interface for macromolecule exchange between a mother and fetus. Revealing symptoms of placenta disease are usually presented later on in the pregnancy with limited treatment options. Access to specialists of placental disease is also limited, particularly in regions distant from urban centres. There is a need for an accessible method to screen patients for risk of placenta disease early. Ultrasound provides a non-invasive imaging modality capable of visualizing the placenta commonly used in obstetric clinics. However, ultrasound images contain unique artifacts that are difficult to interpret with visual inspection. This thesis presents a pipeline of three ultrasound processing methods developed to aid placenta ultrasound analysis.

The first is a method to detect ultrasound acoustic shadow artifacts that obscure anatomy. Radiofrequency signals and pixel entropy were analyzed to identify acoustic shadow in images. A clinical study was performed to obtain ultrasound scans from 37 subjects to evaluate performance. A Dice coefficient of 0.90 ± 0.07 for radiofrequency-based and 0.87 ± 0.08 for pixel entropy-based techniques was achieved when compared to manual shadow detection.

The second is a method to segment the placenta in images preprocessed by shadow detection using a convolutional neural network. Performance was evaluated on data from 1364 fetal ultrasound images from 247 patients. A Dice coefficient of 0.92 ± 0.04 was achieved when compared to manual segmentation.

The third is a method to classify placenta appearance as either normal or abnormal. Images were preprocessed with the first two methods to provide a placenta-only image. A residual convolutional neural network was then used to classify the placenta appearance. Performance was evaluated on 7831 fetal ultrasound im-
ages from 367 patients. Placenta classification achieved a sensitivity of 0.91 and a specificity of 0.87 when compared to classification by physicians.

The methods demonstrate the capability of ultrasound physics analysis and machine learning methods in processing placenta ultrasound images. The results show the potential for developing a tool in the future to assist physicians in analyzing the placenta to screen for disease.
Lay Summary

Several pregnancy disorders originate from the placenta, which is an organ that provides the interface between the mother and fetus. Many placental diseases are difficult to detect until late in the pregnancy, when treatment options are limited. Current practice involves 1-2 ultrasound scans and physician assessment to manually detect placental abnormality. This has challenges such as being limited to the capability of visual analysis, requiring specialized training, and being time-consuming. This thesis addresses the challenges by developing a software tool to automatically classify a placenta appearance as normal or abnormal to aid physicians in screening for placental disease. The tool was developed using a combination of ultrasound physics analysis and artificial intelligence techniques. The tool is comprised of three components: one component to identify regions of signal loss in an image, one component to isolate the placenta, and one component to classify the placenta appearance as normal or abnormal.
Preface

The thesis is derived from two manuscripts. The first manuscript has been published online and is currently in press [41], the second has been accepted to be published in the proceedings of an international conference in July 2019.

Chapter 3 is derived from the published work of the first manuscript:


In the first manuscript, the author designed methods, collected data, and analyzed results of an acoustic shadow detection tool. The author designed and conducted a clinical study to collect all the data involved with technical assistance from Farah Deeba and Rohit Singla. Farah Deeba also contributed to providing a tool for geometric conversion of linear to curvilinear ultrasound images and processing the echo envelope of ultrasound images. The author was the primary contributor to the manuscript with assistance from Rohit Singla. Professor Robert Rohling provided technical guidance, insight, and definition of the problem to be addressed.

The first manuscript involved ultrasound imaging of adult volunteers. All volunteers provided informed written consent. The study was approved by the Clinical Research Ethics Board under the title: Ultrasound Acoustic Shadow Detection (ID: H18-00368).

Chapter 4 is derived from the following work of the second manuscript:

In the second manuscript, the author was the main contributor in designing and conducting a clinical study to obtain clinical ultrasound and patient chart data. The author processed clinical data after Ryan Yan retrieved and filtered the data from the hospital server. The author designed the algorithm for the system and analyzed the results. Delaram Behnami provided suggestions in improving the architecture of the neural network for acoustic shadows. Rohit Singla provided assistance in developing the protocol to the study. The author was the primary contributor to the manuscript, with Ryan Yan and Dr. Chantal Mayer as contributors to the introduction section. Dr. Chantal Mayer providing training in processing the data and clinical insight to the problem. Professor Robert Rohling provided technical insight and definition of the problem to be addressed.

The second manuscript involved collection of retrospective ultrasound and patient clinical chart data. All of the data was de-identified. The data collection was approved by the Clinical Research Ethics Board with the title: Placental Ultrasound Database (ID: H18-01199).

Chapter 5 presents the method, clinical data, and evaluation of a placenta classification method. A manuscript is being prepared for the work in Chapter 5. The author processed the data after Ryan Yan filtered and retrieved the data from a hospital server. The author and Lucas Porto contributed equally to design the software used to address the problem and analyze the data. Ryan Yan and Dr. Chantal Mayer provided clinical insight to pathological details of the problem. Professor Robert Rohling provided technical insight and definition of the problem to be addressed.
# Table of Contents

Abstract ................................................................. iii  
Lay Summary .......................................................... v  
Preface ................................................................. vi  
Table of Contents ................................................... viii  
List of Tables ......................................................... xi  
List of Figures ......................................................... xii  
Glossary ................................................................. xx  
Acknowledgments ..................................................... xxii  

1 Introduction ......................................................... 1  
1.1 Motivation ......................................................... 1  
1.2 Thesis Objectives ............................................... 4  
1.3 Contributions .................................................. 5  
1.4 Outline ......................................................... 7  

2 Background .......................................................... 8  
2.1 Placenta Anatomy and Physiology ............................. 8  
2.2 Placenta Development ......................................... 10  
2.3 Placenta Pathology ............................................. 11
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>Transducer properties for different imaging scenarios.</td>
<td>31</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Mean Dice coefficients for different imaging scenarios ± standard deviation.</td>
<td>41</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>The mean Nakagami $\omega$ and Entropy values of different anatomy, transducer, and shadowing region ± standard deviation. Values are consistent among different transducers and anatomical regions. The variance of entropy and Nakagami $\omega$ in one imaging region and transducer setting is less than the variance across different regions and transducers for shadows and non-shadows.</td>
<td>44</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>The mean Nakagami $\omega$ and Entropy values of all anatomy and transducers for different shadowing regions ± standard deviation.</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Dice similarity coefficients when comparing different detection methods to manual segmentation.</td>
<td>59</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>The recorded sensitivity and specificity of the placenta ultrasound assessment classifier.</td>
<td>72</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>The recorded sensitivity and specificity of the clinical outcome classifier.</td>
<td>72</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>An example of an ex-vivo placenta. An image of the a) fetal surface and b) micrograph of the placenta are shown. Pathological analysis relies on studying the micrograph of the placenta to observe cellular structure. However, this is invasive and requires extraction of the placenta after delivery. This results in difficulty in studying the placenta with histology techniques, motivating ultrasound image studies presented in this thesis. Image taken by Farah Deeba, used with permission.</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>The full placenta ultrasound classification system. The system contains three components, each serving as a precursor to its successive component. The input to the system is a placenta ultrasound image and the output is a binary classification of whether the placenta has normal or abnormal appearance.</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Images of the a) fetal and b) maternal surfaces. The fetal surface is where the umbilical cord is attached and the maternal surface contains lobes of cotyledon with villi that allow diffusion of maternal blood in the intervillous space. Image taken by Farah Deeba, used with permission.</td>
</tr>
</tbody>
</table>
The operation of an ultrasound transducer. The piezoelectric crystal generates an acoustic wave, which propagates into some material or tissue. As the wave propagates, through each tissue, a portion of the wave is transmitted, a portion is backscattered, and a portion is reflected at a boundary. The reflected wave reaches the transducer and vibrates the crystals, generating a voltage. From the time the wave has traveled and the intensity, the processing unit can compute the depth and intensity of the signal. With an array of crystals and continuous reflects at each depth of the tissue, an intensity map of the region being scanned is produced, resembling an ultrasound B-MODE image.

Artist illustration of a fetal ultrasound with major components indicated. A fetal ultrasound is recommended twice by American College of Obstetricians and Gynecologists (ACOG). The illustration displays the placement of an ultrasound transducer, the placenta, and the B-MODE image produced. Image is licensed to Bruce Blaus under a Creative Commons Attribution License (by 4.0): creativecommons.org/licenses/by/4.0/.

A B-MODE ultrasound image of a fetus with anatomy labelled. Note that different anatomy exhibit different texture patterns, with the placenta being a homogeneous, medium-intensity region. The developed bones, such as parts of the skull are recognized by regions of brightness which have darker regions below and the placenta is recognized by an even compressible disc that has lower intensity of the surrounding uterine tissue. This image was obtained from the clinical study ran for the purpose of this thesis (ID: H18-01199).
Figure 2.5  An example of a neural network architecture. The circle represent nodes and lines represents the edges, which data from a previous node is processed as an input to the next node. The input layer is the column of nodes to the left and the output layer is the single column of node to the right. The edges are commonly assigned weights, which modify the input to the successor node. These weights are adjusted with each epoch until an optimized collection of weights configures the neural network to process input data to produce an output with optimized accuracy. The strength of a neural network is the interconnections in the hidden layers, which combine detailed feature computations for complex correlation of features to the output. Note that a neural network does not necessarily have only one input and one output node and can be an array of data.

Figure 2.6  A visualization of a (convolutional neural network (CNN)). A CNN is different from lower dimensional neural networks in that the input is more complex, usually existing as a higher dimensional tensor. The data is processed through neurons with convolutional kernels, which are similar dimensional tensors that convolve with the data from a previous neuron. The advantage of this process is that the output of a convolutional layer represents highly complex feature data, which can be used for detailed correlation to predict a label. A CNN usually contains other utility functions, such as max-pooling to decrease the size of images or densification to combine the width of layers.

Figure 3.1  Example of an acoustic shadow in an arm ultrasound. An acoustic shadow is created at bone-tissue boundary of the radial joint, radius, and ulna, resulting in a bright region followed by an almost total signal loss further down the scanline. As a result, anatomy beyond the shadow boundary is obscured.
Figure 3.2  Processing steps for RF and B-MODE shadow detection. RF processing is used if RF data is available and involves fitting the Nakagami distribution onto the echo envelope of each RF scanline before adaptive thresholding with Otsu’s method. In many cases, there may only be access to B-MODE image data, for which an entropy map is computed and similar adaptive thresholding is used to detect shadows.

Figure 3.3  Visualization of the three scanning locations in the clinical study including the a) forearm, b) elbow, and c) ribcage with corresponding images in d), e), and f). Acoustic shadows are expected due to the bone-tissue boundary in the arm and ribs.

Figure 3.4  A comparison of a scan of the forearm from a linear transducer with the a) brightness-mode and b) logarithmic echo envelope of the RF data. The B-MODE data is visually enhanced to remove the speckle for a smoother image, but the speckle in the RF data can be leveraged to understand the scattering interactions of the acoustic wave in certain materials.

Figure 3.5  The processing method for RF data with a) the raw RF data with a sample scanline highlighted, b) the echo envelope of the raw RF data scanline, c) the echo envelope of the entire RF image, and d) the logarithmic scaled echo envelope image. This processing is required for the RF data to contain statistically distinguishable characteristics between patches to characterize acoustic shadows.
Figure 3.6  A visualization of the B-mode and RF parameter maps. The b) Entropy Map was computed from processing of the a) original B-MODE image and the d) Nakagami $\omega$ map was computed from the c) echo envelope. Note that the echo envelope contains noticeable speckle, which has been used to fit a Nakagami distribution to characterize shadow. The region at depth 2.50 cm and scanlines 32-40 is attenuation and not a shadow. This is an important distinction in shadow detection and both maps show the region as below a threshold to flag a shadow boundary.

Figure 3.7  Creating the scanline paths on a curvilinear ultrasound image of the right ribcage and the forearm with different depth settings on the curvilinear transducers. Two initial paths are creating by computing the slopes of the two lateral edges, which are detected by tracking the ring-down artifact. The remaining paths are creating by linearly interpolating the slope of pixels in between the lateral edges.

Figure 3.8  Images of both RF and B-MODE shadow detection performed on a gelatin phantom with two wooden slits embedded at a depth of 0.75cm and 2.50cm. The phantom was made to simulate shallow, deep, and non-shadow regions. The methods were capable of shadow detection with a Dice coefficient $> 0.95$, though noticeable errors were present at high-gain images for the B-mode method. This is expected as B-MODE methods rely on pixel gray level, which may vary due to operator settings.

Figure 3.9  A comparison of the original B-mode images, the detected shadows manual detection, RF detection, and B-mode detection. Both detection methods perform similarly to manual detection. Both methods perform slightly less accurately on curvilinear images, potentially due to the reduced resolution from interpolating the scanlines. Most errors of RF detection occur near the shadow boundary, potentially due to the transitioning speckle from non-shadow to shadow.
Figure 3.10  Histograms of Nakagami parameters and entropy values in shadow and non-shadowing regions. The Nakagami $\omega$ and Entropy distributions have a more noticeable delineation between shadowing and non-shadowing distributions compared to the Nakagami $m$ parameter, which was not used to threshold shadow boundaries. Entropy is very minimal is continuous dark shadow regions, which is expected due to the minimal variations in pixel gray level.

Figure 4.1  Pipeline of the placenta detection method. After preprocessing an input image, the image is fed to two parallel paths. The first is through convolutional layers similar to U-Net to produce a preliminary placenta segmentation. The second is an automated shadow detection method to tune the optimization weights of the preliminary segmentation. After convergence of the optimization, a final output placenta map is produced.

Figure 4.2  Visualization of the convolutional layers. The input image is convolved with a series of weighted activation feature maps and the resulting activation output is compared to the ground truth to optimize intermediate weights such that the output of an activation pathway corresponds to either placenta or non-placenta. In this particular architecture, the feature maps are down-scaled with max-pooling to reduce computational complexity with the more detailed maps before being upscaled again. Several hyperparameters and dropout pathways were adjusted to improve placenta segmentation.

Figure 4.3  Output segmentation maps compared with manual segmentation. Three cases are presented including one of a placenta in plain view (a,d,g), one of a placenta in partial view (b,e,h), and one of a placenta with an acoustic shadow (c,f,i). The focus of the method is to handle cases with acoustic shadows, where the detection is able to identify placenta regions even though they are not a single closed contour.
Figure 4.4  Output segmentation maps of the lowest accuracy images. Case 1 (a,d,g) missed identifying placental regions in the bottom left, although it was able to identify the center dark lesion as part of the placenta. Case 2 (b,e,h) misidentified fetal parts on the bottom as placental regions. Case 3 (c,f,i) misidentified upper uterine wall and lower fetal regions as part of the placenta. In cases of false positives, the algorithm may improved with additional training segmentations of the same dataset except with different anatomy identified, such as the uterine wall of fetal limbs. .......................... 60

Figure 5.1  A visualization of the ResNet architecture. Only two residual blocks are displayed for clarity. The full architecture used contained 50 layers. ResNet is a much deeper network than U-Net used in placenta segmentation, allowing more complex details to be activated. The residual blocks provided references of previous layers to initialize the output of a deeper layer to assist optimization. .......................... 70

Figure 5.2  The confusion matrix of the classifier to predict if a placenta had been assessed as normal or abnormal appearance from the ultrasound image using the ultrasound assessment classifier. .......................... 73

Figure 5.3  The confusion matrix of the classifier to predict if a placenta had been assessed as being associated with disease or no disease using the clinical outcome classifier. .......................... 74
Figure 5.4  The class activation maps of a) normal placenta and a b) abnormal placenta from the ultrasound assessment classifier. The activation maps show the segments of the placenta that provided the most important features used to discriminate a placenta as normal or abnormal. Visually, there is no recognizable pattern with the location of the activations. This indicates that the activations are not location-specific and that features throughout the entire placenta was used to correlate with an abnormal or normal prediction. This may include details such as small lesions or inhomogeneities in echotexture that are not visible to the human eye, unless magnified. This map may aid physicians in identifying regions of interest for pathological study of the placenta.

Figure 5.5  The output of the first 6 convolutional kernels. From the output, a physician can visually observe what sections and what type of features are being extracted from the image, although it is difficult to quantify without detailed numerical analysis.
Glossary

2D  two-dimensional

3D  three-dimensional

ACOG  American College of Obstetricians and Gynecologists

ADAM  adaptive momentum estimation

AI  artificial intelligence

B-MODE  brightness-mode

BCW  British Columbia Women’s Hospital and Health Centre

CAM  class activation map

CNN  convolutional neural network

CT  computed tomography

EMMA  Evaluating Maternal and Fetal Markers of Adverse Placental Outcomes Clinic

GPU  graphics processing unit

ILSVRC  ImageNet Large Scale Visual Recognition Challenge

IUGR  intra-uterine growth restriction

ML  machine learning
**MRI** magnetic resonance imaging

**RF** radio-frequency
Acknowledgments

I sincerely thank Professor Robert Rohling for his role as my graduate studies supervisor. Your insight in ultrasound imaging and research has provided a most significant period of growth in my academic capability in the last two years. Your guidance was essential for me to conduct research in a professional manner. I would also thank Dr. Chantal Mayer for her role as my clinical supervisor. The insight to the clinical challenges of ultrasound imaging was invaluable and I am truly appreciative of the time you took from your schedule to guide me through my research. Both supervisors were instrumental to providing me with an understanding of how to advance the fields of engineering and medicine to define the work I am pursuing in the future.

I thank the uniquely amazing colleagues I have had the fortune of working with through the years. Specifically, I thank Rohit Singla for guiding me with his knowledge of engineering research with clinical applications. I thank Farah Deeba for teaching me essential methods to operate and process ultrasound machines. I thank Victoria Lessoway for your enthusiastic diligence to assist me in understanding ultrasound with your unparalleled experience as a sonographer. I thank Lucas Porto for providing his artificial intelligence knowledge to work in tandem on difficult artificial intelligence problems.

My experience in the Robotics and Control Lab has convinced me that there is no better lab for me to have worked at. Every single member has been kind, helpful, and skillful and I am lucky to learn a wealth of knowledge from the top ultrasound and artificial intelligence experts in the world.

I thank the Natural Science and Engineering Research Council (NSERC) and Canadian Institutes of Health Research (CIHR) for funding this work.
I thank my friends for the support through my graduate studies, particularly those who have taken the time to help me in writing this thesis. Thank you Jessica Lo for your daily acts of encouragement to make each day enjoyable. Thank you Ranesh Saha, Martin Yu, Erick Leung, Jason Wong for providing me a reprieve which I can retreat at any time to find solace from the fray.

Finally, there is no group that has spent more voluntarily hours in supporting me than my family. I thank my mom and dad, Yvonne and Tim, who work tirelessly without ever a complaint to support me at my every need. Thank you for waking early or staying late to cook for me without a second thought. Thank you for the brave jump from the familiarity of your hometown villages to a new life in Canada thirty years ago to provide your future child with opportunities you did not have by working days, nights, and holidays to put food on the table. My goal in life is to hope that I come close to repaying you for all your sacrifices. None of my achievements or awards are without these two hidden figures. No man is an island, but rather, I am kept above water by the love and support of my parents.
Chapter 1

Introduction

1.1 Motivation

During pregnancy, the body of the mother is subjected to a radical period of change. The development of an offspring from an embryo to a newborn fetus is coupled with a change in metabolism, development of new organs, and complex molecular interactions between the mother and the fetus [89]. The intricate environment required for healthy fetal development is thus a sensitive process that is susceptible to many potential health complications [74]. One essential component of this environment to pregnancy is the placenta. The placenta is an organ that develops during pregnancy that functions as an interface for macromolecule exchange between the mother and the fetus [14]. The placenta and an example micrograph is shown in Figure 1.1.

The placenta utilizes its anatomy to allow the exchange of nutrients, antibodies, and waste between the mother and the fetus, as discussed in further detail in Chapter 2. The growth of the fetus is reliant on a functioning placenta for the exchange of essential molecules [79]. Thus, malfunction in the placenta can impede the healthy development of the fetus.

There are several challenges in detecting and treating diseases associated with the placenta. First, the symptoms of placental disease are difficult to detect early. As the placenta is a constantly developing organ there is a rapid change of how the organ interacts with the mother. This results in difficulty in observing consis-
Figure 1.1: An example of an ex-vivo placenta. An image of the a) fetal surface and b) micrograph of the placenta are shown. Pathological analysis relies on studying the micrograph of the placenta to observe cellular structure. However, this is invasive and requires extraction of the placenta after delivery. This results in difficulty in studying the placenta with histology techniques, motivating ultrasound image studies presented in this thesis. Image taken by Farah Deeba, used with permission.

tent symptoms before the placenta is fully developed and many symptoms unique to placental disease present later in the pregnancy. For instance, in preeclampsia, where there is poor development of vasculature between the mother and the fetus, the identifying symptoms are protein in urine and hypertension. Protein in the urine does not usually occur until after the 20th week of pregnancy [75] and hypertension is a symptom shared with many other disorders. The treatment methods are limited, with delivery being the most effective treatment of the disease. Before delivery, the patient is recommended bedrest and are treated for symptoms such as hypertension to prevent further complications. Early treatment options are rare. Low-dose aspirin [92] and reductase inhibitors [24] has shown efficacy in randomized controlled trials, but only when administered early in the pregnancy. Hence, early detection methods are necessary for effective treatment options.

Second, it is difficult to track the progression of disease and how disease affects the placenta. In addition to the first challenge of diagnosing placental disease early, it is also difficult to monitor how the placenta changes. Diagnostic options to monitor the placenta, such as a biopsy, is invasive and requires laboratory resources. The invasive nature of biopsies means it may not be feasible for valuable
longitudinal studies.

Third, there is a challenge with access to diagnostic resources for placental disease. Detailed analysis of structural abnormalities, such as lesions and placental lakes, is difficult and commonly require specialized training in maternal-fetal medicine. Thus, patients who reside away from specialized centers have difficulty accessing care. The lack of accessible care results in potential progression and delayed of undiagnosed diseases, leading to adverse outcomes for the mother and fetus.

To detect placenta abnormalities, ultrasound imaging is a promising tool to assist analysis. Ultrasound is already a popular imaging modality for pregnancies as it provides unique advantages over other tools. Ultrasound imaging contains no ionizing radiation, unlike X-rays or CT scans. Ultrasound imaging is also significantly cheaper and more portable than MRI scans, with handheld ultrasound models available in the market [65]. With its existing popularity, there is a motivation for ultrasound imaging to be developed to detect placenta abnormalities early in the pregnancy and track the progression of placental disease throughout the pregnancy.

However, ultrasound imaging contains its own challenges for placental analysis. A key challenge is that analyzing ultrasound images visually is difficult due to artifacts arising from the interaction of acoustic waves with matter. Ultrasound contains unique artifacts such as acoustic shadows, where certain boundaries reflect the majority of a signal, resulting in dark regions that obscure any anatomy. Ultrasound images also contain speckle, where acoustic waves scatter throughout matter, creating granular patches that are difficult to consistently recognize. The texture of ultrasound images vary depending on the model of the machine, transducer geometry, and operator settings such as depth and frequency. The current practice at many hospitals is that a trained sonographer carefully scans a patient for an angle that captures the placenta. A physician with specialized training in analyzing fetal ultrasound will then observe the placenta. Current observation methods are limited to subjective metrics, such as if there is abnormal echotexture or if the placenta is irregularly shaped [60]. However, these observations are qualitative and are highly subjective to the observer, resulting in difficulty to record standardized data.
There is thus a strong motivation to advance ultrasound imaging, which is increasing affordable and portable, to assist physicians in analyzing placenta to predict disease. There are several requirements for the solution. First, the solution must be capable of completing a task with accuracy similar to manual analysis, whether it is identifying the location of the placenta or classifying a placenta as abnormal. Second, the ultrasound images must be processed in real time for large-scale data applications in an automatic way. Third, the solution must also be usable with minimum training, such that no expert setup is required. Fourth, the solution must also be capable of handling different types of images from different machines and operator as to not be overtuned for one specific scenario.

1.2 Thesis Objectives

This thesis ultimately presents a pipeline to classify a placenta as normal or abnormal to aid the analysis of placental health. This work aims to provide a tool that can be used alongside standard practice to aid the screening of placental disease. Although the work presented does not aim to replace final diagnostic procedures, it aims to provide physicians with an additional tool that is automated and produces quantified placenta appearance classification data. This may produce benefits such as additional data for diagnostic decisions and providing an option to analyze the placenta appearance when a combination of a sonographer and radiologist are unavailable, such as in rural areas.

This thesis proposes three methods that compose the pipeline. Each method is a precursor to the successive method with the objective of addressing challenges existing in developing a full end-to-end placenta classification system. The first method detects an acoustic shadow of an ultrasound image, aiming to address the issue that acoustic shadows are unique ultrasound artifacts that may confound artificial intelligence pattern recognition. The second method automatically segments a placenta from an acoustic-shadow-filtered ultrasound image, aiming to address the issue that a placenta need to be automatically extracted before further analysis. The final method classifies a placenta segmentation as having normal or abnormal appearance.
1.3 Contributions

The full end-to-end placenta ultrasound classification tool is illustrated in Figure 1.2. The system is separated into three methods and each contribution corresponds to a block in Figure 1.2.

1. A method was developed for the automatic identification of acoustic shadows, given the cell resolution size of an ultrasound transducer. This work proposes two approaches to analyze ultrasound images with shadows. One approach utilizes radio-frequency (RF) data to analyze statistical distributions of ultrasound speckle. The distribution is then characterized as either shadow or non-shadow. The second approach utilizes brightness-mode (B-MODE) data to compute the entropy of scanlines. The shadow boundary is then detected using adaptive thresholding of the entropy levels. This work has been tested on data collected from 37 volunteers and 222 images and results have been evaluated against manual detection of acoustic shadows. The acoustic shadow detection method serves as a pre-processing input to the second contribution.

2. A method was developed for the automatic segmentation of placenta in a fetal ultrasound image. This work proposes the usage of a cascaded convolutional neural network (CNN) for automatic classification of an ultrasound region as placenta or non-placenta texture, using acoustic shadowing as a attention-modifier to ignore regions of signal loss. This method generates segmentations in automatically and has been tested on data collected from 247 patients and 1364 fetal ultrasound images. This work has been evaluated against manual segmentations of the placenta. The placenta segmentation method serves as a pre-processing input to the third contribution.

3. A method developed for the automatic classification of a segmented placenta ultrasound image as having either normal or abnormal appearance. This work proposes a residual neural network for automatic classification of a placenta as either abnormal or normal appearance. This method generates classifications in real-time and has been tested on data collected from 7831 fetal ultrasound images from 367 patients.
Figure 1.2: The full placenta ultrasound classification system. The system contains three components, each serving as a precursor to its successive component. The input to the system is a placenta ultrasound image and the output is a binary classification of whether the placenta has normal or abnormal appearance.
1.4 Outline

This thesis presents the physiological background of the placenta relevant to placental disease analysis and the details of the three systems developed. The evaluation of the three systems is presented with a discussion of the results, including future implications and limitations. The outline of the thesis is as follows:

- **Chapter 2: Background** Presents relevant placenta physiological, clinical practice of placenta analysis, basic US theory, and topics in machine learning used to develop the methods in this thesis.

- **Chapter 3: Acoustic Shadow Detection** Presents the clinical study to capture acoustic shadows, the system developed to detect shadows, and the evaluation of the system.

- **Chapter 4: Automatic Placenta Segmentation** Presents the clinical study to obtain ultrasound images of the placenta, the system developed to segment the placenta, how the acoustic shadow detection work serves as a precursor, and the evaluation of the system.

- **Chapter 5: Automated Placenta Classification** Presents the clinical study to obtain ultrasound images of the placenta with labeled classification, the system developed to classify the placenta, how the placenta segmentation work serves as a precursor, and the evaluation of the system.

- **Chapter 6: Conclusion and Future Work** Summarizes the work completed in this thesis, contributions, implications of the results, limitations, and potential developments in the future to address the challenges of placenta analysis.
Chapter 2

Background

This chapter discusses the anatomy and physiology of the placenta with the purpose of highlighting important pathological details for placenta analysis. The fundamentals of ultrasound imaging for the placenta are then presented. The standard of practice in interpreting placental ultrasound images and challenges are then discussed to highlight the gaps in current placenta images. From the existing needs of the current practice, this chapter finally the fundamentals of machine learning (ML) and how it may address placenta imaging challenges.

2.1 Placenta Anatomy and Physiology

The anatomy of the placenta contain structures essential to the exchange of macromolecules between the mother and the fetus. Thus, changes to the structure discussed can lead to abnormal development. The appearance changes of the placenta during disease is an active area of research [15] and this thesis aims to develop a tool to assist in monitoring the appearance of the placenta with ultrasound imaging.

The placenta is an organ that develops only during pregnancy. It is visually shaped like a compressible disc, averaging 20 cm in length and 2-4 cm in thickness [85]. One surface of the placenta contains blood vessels attached to maternal circulation and the opposite surface contains the umbilical cord which is connecting to the fetus. The maternal and fetal surfaces of the placenta are displayed in Figure 2.1.
Before pregnancy, the maternal surface is composed of endometrial lining connected to the uterus. The endometrial lining changes 6-8 days after fertilization, developing into the decidua, or uterine lining, of the placenta [94]. The maternal surface contains cotyledon, which are lobes that are separated by channels called sulci. The cotyledons contain villi, which are small projections containing placental capillaries that are exposed to the intervillous space [77]. The maternal veins and arteries connect to intervillous space, providing a flow of blood from the mother to the placental capillaries and to the fetal surface [43].

The fetal surface contains the chorion, a thick membrane in the interior of the placenta that contains the fetal capillaries connected to the intervillous spaces. The chorion is covered by the amnion, a membrane that produces amniotic fluid, which provides a compressible barrier between the fetus and uterus to dampen movement and protects the fetus. The centre of the placenta contains the umbilical cord, which serves as a conduit to the fetus from the placenta [43].

The anatomy of the placenta allows for several functions. As the fetus has not developed lungs early in the pregnancy, the fetus cannot generate usable energy through cellular respiration on its own. In addition, the fetus has no self-sustaining source of nutrients such as glucose and proteins. Instead, blood containing oxygen and nutrients from the mother travels through the intervillous space, across the placenta, and through the umbilical vein to diffuse into the fetal bloodstream.

Figure 2.1: Images of the a) fetal and b) maternal surfaces. The fetal surface is where the umbilical cord is attached and the maternal surface contains lobes of cotyledon with villi that allow diffusion of maternal blood in the intervillous space. Image taken by Farah Deeba, used with permission.
Conversely, waste such as carbon dioxide and urea are produced from the fetus, which flows through the umbilical arteries to diffuse through the intervillous space to maternal circulation [35].

The placenta also has an endocrine function to secrete hormones for the fetus to maintain homeostasis [63]. The placenta secretes human chorionic gonadotropin to signal the surrounding follicular tissue to secrete progesterone. Progesterone then promotes the growth of the uterine wall. Further on in the pregnancy, the placenta is able to promote progesterone on its own. Human placental lactogen is secreted to reduce the mother’s usage of glucose such that more glucose is available for the fetus to grow [66]. The placenta also secretes estrogen, which increases blood flow to the fetus. During labor, estrogen stimulates the secretion of prolactin to produce milk from the mammary glands [66].

Finally, the placenta provides immune protection for the fetus through two methods. First, the placenta serves as a physical barrier between the mother and fetus. Since the fetus contains genetic information from two sources, there is a possibility that the maternal immune system identifies the fetus as a foreign body and initiates an immune response [27]. However, the placenta prevents direct contact between the maternal and fetal circulation as it only allows circulation through diffusion in the intervillous space [27]. Second, the placenta allows for the transfer of antibodies through the same channel as the transfer of nutrients [21]. As the immune system of the fetus is developed, the antibodies from the mother can provide additional protection to the fetus once it is exposed to the external environment after birth.

Abnormalities in the placenta can then lead to the dysfunction of its physiological function where the fetus does not receive nutrients, hormonal stability, and immune protection.

2.2 Placenta Development

The development of the placenta results in rapid changes to the observable structures throughout the pregnancy as it is an organ that is delivered and removed at birth. Hence, the rapid development of this organ results in difficulty studying anatomical changes during disease. This challenge is relevant to the objective of
the thesis, which aims to develop a tool that can monitor changes in placenta using ultrasound imaging.

The placenta begins growing roughly four days after fertilization. The placenta develops from differentiating cells from the zygote. The outer cell layer, called the trophoblast cells, develop into the placenta, which will be implanted to the uterine wall. The trophoblast then develop projections that penetrate to the maternal veins and arteries. The projections develop into chorionic villi, where the gas and waste exchange occurs. In the following weeks, the amnion develops from the inner cell mass and envelopes the fetus [26]. The development of chorionic villi and amnion are essential events in placenta development and malfunctions in this process may lead to complications relevant to placental disease such as poor fetal circulation throughout the pregnancy.

2.3 Placenta Pathology

Two diseases are of particular interest as there are numerous studies linking the disease to placenta development. Although many other diseases associate with the placenta exist, the diseases discussed in this section were selected as they are more likely to be associated with structural changes with the placenta, which may be observable in ultrasound [8] [25].

The first disease is known as preeclampsia. Preeclampsia is a disease associated with poor maternal-fetal exchange, affecting 1-3% [80] of all pregnancies and is a major factor in one third of obstetric morbidity [98]. The disease is characterized by clinical symptoms of hypertension and proteinuria, arising after the 20th week of pregnancy. Preeclampsia can progress to maternal complications including seizures, long-term renal failure, long-term cardiovascular disease, and hypertensive encephalopathy [53]. Fetal complications may include perinatal death.

There is no definitive cure for preeclampsia and the most effective treatment is the delivery of the baby. However, this is only possible in late pregnancy. Before delivery is possible, the patient may be prescribed bed rest, increased monitoring of blood pressure, and low-dose aspirin. Aspirin has been hypothesized to inhibit the synthesis of thromboxane A₂ [19], which is a vasoconstrictor that promotes aggregation of platelets and resulting hypertension. Past meta-analyses of controlled
trials with low-dose aspirin only show efficacy when administered early [78]. Early detection monitoring is thus required to assess the health of the mother accurately and respond quickly to symptoms.

The cause of preeclampsia is suspected to be poor placentation [32]. Poor placentation starts as early as 6 weeks into pregnancy when the trophoblast layer is unable to fully rupture and access the maternal capillaries [73]. Chorionic villi then cannot establish a strong fetal-maternal interface and circulation is hindered, leading to fetal hypoxia and oxidative stress. An increasingly hypoxic placenta causes clinical maternal symptoms of hypertension, proteinuria, and vascular dysfunction [76]. The microstructure of the placenta has been observed to change in preeclampsia. Preeclampsia placenta was observed to have increased glycosaminoglycans, which increases tissue viscosity [86] and may decrease the extent a tissue can be deformed. Placentas from poor placentation were observed to have reduced tissue to fluid ratio [61] as a result of smaller chorionic villi and more intervillous space. Placentas from poor placentation were also observed to have a lower collagen to elastin ratio compared to normal placenta [22], which may lower structural strength. However, it is not known how poor placentation leads to lower proliferation of chorionic villi and whether smaller chorionic villi is specific to preeclampsia. Relevant to ultrasound imaging, it is also not fully known how changes in the chorionic villa of the placenta appear on an ultrasound image. This is a major challenge in tracing changes in the placenta as histological observation requires continuous biopsy of the placenta, which is invasive.

The second disease of interest is intra-uterine growth restriction (IUGR). IUGR occurs in 5-8% in all pregnancies when the fetus undergrowth poor growth and its weight is under the 90th percentile [31]. There are many causes of IUGR, including such as malnutrition, existing disease, lack of oxygen at high altitudes, and preeclampsia [93]. Abnormalities in the placenta are also suspected to be a cause of IUGR as a small or poorly developed placenta may not provide the fetus with sufficient nutrients to grow. IUGR can lead to a variety of complications, including neurodevelopment disorders, hypothermia of the fetus, and hypoglycemia at birth [9]. Due to the wide range of causes of IUGR, treatment options are limited. Bed rest is recommended, the mother is closely monitored and repeated ultrasounds and blood tests are taken. If preeclampsia is suspected to be the cause of IUGR, then
low-dose aspirin is prescribed. Similar to preeclampsia, past meta-analyses show that aspirin is only effective at reducing symptoms of IUGR when administered early in the pregnancy [11]. Hence, there is a similar motivation to detect IUGR early to provide more effective treatment options and monitoring of symptoms.

In both cases, there is a potential relationship between the progression of the disease and changes in the placenta. Hence, an improved method of ultrasound imaging may help monitor subtle texture changes in the image that may be associated with structural changes due to preeclampsia or IUGR.

### 2.4 Fetal Ultrasound Imaging

Ultrasound imaging is the sole modality investigated in this thesis as it is a relatively inexpensive, accessible, and non-invasive imaging method compared to other modalities. It is the main modality used in obstetrics.

This section discusses the basic principles of ultrasound imaging and how ultrasound imaging is used in practice for monitoring pregnancies. The gaps in current ultrasound imaging practice are discussed to motivate the objectives of this thesis.

#### 2.4.1 Introduction to Ultrasound

Ultrasound is an imaging modality that is a non-invasive method to visualize anatomy. Ultrasound does not produce ionizing radiation as opposite to X-ray or computed tomography (CT) scans. Ultrasound is also faster and significantly less expensive than CT, and magnetic resonance imaging (MRI). Recently, ultrasound devices are being developed to be increasingly compact, such as the development of portable ultrasound devices [65]. Thus, ultrasound imaging is the most common modality to image the fetus, as the ionizing radiation may damage the fetus. In addition, limited availability of MRI cannot handle the large volume of pregnancies [12].

Ultrasound operates on the principle of how acoustic waves interact with matter. An ultrasound system is commonly composed of a transducer, a processing unit, and an user interface (UI). The basic operation of the wave propagation and detection by a transducer is illustrated in Figure 2.2 and an illustration of a fetal ultrasound is shown in Figure 2.3. The process of which an ultrasound image is produced is as follows:
Figure 2.2: The operation of an ultrasound transducer. The piezoelectric crystal generates an acoustic wave, which propagates into some material or tissue. As the wave propagates, through each tissue, a portion of the wave is transmitted, a portion is backscattered, and a portion is reflected at a boundary. The reflected wave reaches the transducer and vibrates the crystals, generating a voltage. From the time the wave has traveled and the intensity, the processing unit can compute the depth and intensity of the signal. With an array of crystals and continuous reflects at each depth of the tissue, an intensity map of the region being scanned is produced, resembling an ultrasound B-MODE image.
1. The ultrasound transducer is placed on the surface of the patient, with the face containing piezoelectric crystals in contact with the patient.

2. Power is supplied to the transducer, which causes the piezoelectric crystals to vibrate, propagating an acoustic wave into the patient.

3. As the acoustic waves propagate through tissue, the waves interact with tissue and the wave is scattered. A portion of the wave is continuously backscattered towards the transducer. As the wave reaches a boundary, a portion of the initial wave reflects back to the transducer, with continuous scattering along its propagation path.

4. The reflected and backscattered waves (echoes) reaching the transducer vibrate the piezoelectric crystals, which generate a voltage.

5. The voltage is measured by the processing unit. The time that the acoustic wave travels is used to compute the depth of the tissue that created that portion of the echo. The magnitude of the voltage is used to compute the intensity of the echo.

6. As there are multiple piezoelectric crystals spread across the transducer, the processor receives intensity information from the length of the transducer. As the acoustic waves are reflected continuously as the wave propagates deeper, the processor receives intensity information depthwise. A two-dimensional map of the imaged region is then computed from the collected set of echo signals after time-delay based beamforming.

7. The processed map is displayed on the UI in grayscale, with darker regions representing low intensity signals and brighter regions representing high intensity signals. This is known as a B-MODE image.

Different anatomy reflects the acoustic waves differently. For instance, a tissue-bone boundary reflects most of the acoustic wave, resulting in a bright region on a B-MODE image. The shape and brightness of regions in B-MODE image can then be interpreted as different anatomical features. There also exists many unique artifacts in ultrasound imaging due to the physics of acoustic waves interacting with
Figure 2.3: Artist illustration of a fetal ultrasound with major components indicated. A fetal ultrasound is recommended twice by American College of Obstetricians and Gynecologists (ACOG). The illustration displays the placement of an ultrasound transducer, the placenta, and the B-MODE image produced. Image is licensed to Bruce Blaus under a Creative Commons Attribution License (by 4.0): creativecommons.org/licenses/by/4.0/.

In addition, ultrasound imaging is known for its grainy visual appearance due to the scattering of acoustic waves, known as acoustic speckle [13]. These artifacts and acoustic speckle result in difficult in interpreting some ultrasound images, as the image does not have as clear anatomical boundaries as CT or MRI. At the same time, certain anatomy are known to exhibit artifacts, such as a tissue-bone boundary reflecting most of the acoustic wave, resulting in a bright horizontal line followed by a dark region below. These artifacts can be leveraged to identify anatomical features, which are useful in preprocessing ultrasound images. Tech-
Figure 2.4: A B-MODE ultrasound image of a fetus with anatomy labelled. Note that different anatomy exhibit different texture patterns, with the placenta being a homogeneous, medium-intensity region. The developed bones, such as parts of the skull are recognized by regions of brightness which have darker regions below and the placenta is recognized by an even compressible disc that has lower intensity of the surrounding uterine tissue. This image was obtained from the clinical study ran for the purpose of this thesis (ID: H18-01199).

Guidelines for fetal ultrasound imaging varies between different jurisdictions. Common guidelines from ACOG (American College of Obstetricians and Gynecologists) recommend two ultrasound examinations: one in the first trimester and one...

2.4.2 Ultrasound in Fetal Imaging

Guidelines for fetal ultrasound imaging varies between different jurisdictions. Common guidelines from ACOG (American College of Obstetricians and Gynecologists) recommend two ultrasound examinations: one in the first trimester and one
in the second trimester or later [60].

The first trimester ultrasound is performed before 14 weeks. It evaluates the presence of the fetus, cardiac activity, and location of early structures such as the gestational sac. The placenta may not have fully completed development at this stage.

The second trimester ultrasound is performed between 18-20 weeks of gestation [60]. At this stage, a physician observes the fetal heart rhythm and estimated the amount of amniotic fluid. An anatomical survey is taken, measuring the presence and size of organs such as the kidneys, stomach, brain, and spine. The uterus and cervix are also examined. At this stage, the placenta should be more visible and is measured. The placenta location and cord insertion site is observed.

The analysis of the placenta currently occurs mainly in the second trimester ultrasound [60]. Several observations are made to determine if the placenta is abnormal:

- **Placenta Size:** The placenta is expected to grow during a pregnancy to a thickness of 2-4 cm and a length of roughly 20 cm. An excessively thick placenta may be indicative of an infection, particularly if the placenta has visible cysts [55]. A significantly small placenta may be related to preeclampsia or IUGR due to poor maternal-fetal circulation and oxidative stress [44].

- **Placenta Location:** The location is defined as relative to the centre of the uterus. Of particular concern is when a placenta is low-lying and exists within 2 cm of the internal cervical opening, called placenta previa [54]. This may cause severe bleeding and a caesarean section is required to deliver the baby.

- **Placenta Cord Location:** The location which the umbilical cord is inserted in the placenta is observed. A normal umbilical cord is embedded near the centre of the placenta. If the umbilical cord is less than 0.5 cm from the edge, the cord may insert through the membrane between the amnion and chorion, known as a velamentous cord insertion [49]. This may cause vessels to rupture and cause an abnormal fetal heart rate. This has also been associated with low birth weight and a poorly developed fetus [48].
• **Cysts and Lesions**: The texture of normal placenta tissue is usually homogeneous. Thin-walled cysts containing liquid can be observed as darker regions, as the liquid does not reflect acoustic waves as strongly as tissue. Cysts are mostly benign, though cysts larger than 4.5cm have been reported to be associated with IUGR [1]. Other lesions with abnormal or inconsistent echotexture compared to the regular homogeneous placenta tissue may be benign growths or tumors. The most common tumor occurs as chorioangiomas, which is observed in 1% of pregnancies. [2]. Chorioangiomas are associated with increased fetal morbidity and mortality, with complications of fetal anemia and potential cardiac failure [103].

2.4.3 **Existing Work and Limitations to Placenta Ultrasound Analysis**

Aside from measurement of placenta size and location, analysis of its appearance in an ultrasound image is subjectively defined. The current guidelines recommend observing regions of bright, dark, homogeneous, and inhomogeneous texture [60]. This thesis attempts to address the rough classification categories by providing more quantitative analyses of placenta appearance, which may aid in more detailed placenta analysis. There is limited work on computational placenta ultrasound analysis methods. They are limited to detecting the location of the placenta and placenta maturity grading.

Placenta detection methods involve segmenting the placenta from an ultrasound image. Stevenson et al. presented a semi-automatic method for placenta segmentation with accuracy similar to manual segmentation [91], requiring manual initialization of the algorithm. [59] using existing artificial intelligence (AI) software to segment the placenta. The method was evaluated on three-dimensional (3D) ultrasound images, obtaining a Dice coefficient of 0.73. However, this was tested on 240 images from the same transducer model and machine. Both methods sought to characterize the texture of the placenta in a quantitative manner to improve the consistency of placenta detection. However, both methods did not address all the challenges of existing placenta analysis in that an automated method is required to be tested on a wide variety of images. Improvements on these methods
are discussed in Chapter 4.

Li et al. [56] developed a model for automated grading of placenta maturity, achieving a sensitivity of 98.04% and a specificity of 93.75% when tested with 443 placenta images. The goal of that study was to quantify the stage of a placenta. This is useful to detect if a placenta is overdeveloped or underdeveloped for its gestational age, which can identify certain developmental disorders [87]. The method used analyzed regions of the placenta with higher order statistics. In a similar way to placenta detection, the authors aimed to characterize the placenta quantitatively for consistent analysis. However, the study was only performed on a limited set of images specifically obtained for maturity analysis and there are limitations in generalizing the method to clinical use. Limitations of that study is discussed in further detail in Chapter 5.

In the current standard of practice, usage of ultrasound is limited to subjective analysis by a trained physician. In past studies for computational methods in placenta ultrasound, the studies do not address the challenges of full automation or being validated over different machines and images. The gaps in the standard of practice and existing studies motivate the work in this thesis, where a combination of ultrasound physics analysis and ML methods were develop for classifying abnormalities in a placenta ultrasound image.

2.5 Machine Learning in Ultrasound

To improve analysis of ultrasound images, ML has shown potential in providing automated recognition of image features. This section discusses selected ML topics relevant to the challenges in fetal ultrasound imaging: image segmentation and image classification. Detailed discussion of specific techniques used is presented in Chapter 4 and Chapter 5.

2.5.1 Basic Principles of Machine Learning

ML techniques are based on the principle of providing data repeatedly to some parameterized model, with each pass of the dataset known as an epoch. With each successive epoch, the parameters of the model are adjusted to maximize the accuracy of the output. In a class of ML methods called supervised learning, the each
data point is labeled with some characteristic, known as the ground truth. The entire dataset is split into two subsets: a training set and a validation set. For the training set, each data point is passed through the model as an input and the model predicts a label. The predicted label is compared with the ground truth to measure the accuracy. This is repeated, with each the parameters in the model adjusted in each epoch until the accuracy increases and converges, resulting in an optimized model. The validation set is then passed through the model as an input and the predicted labels of the optimized model are compared to the ground truth. The difference in the validation set is that the model does not re-process the validation data as the validation accuracy is meant to be the most representative test of the model’s ability to be generalizable. The methods discussed in this thesis will all use supervised learning.

2.5.2 Neural Networks and Deep Learning

Deep learning is a category of ML methods that involves multiple intermediate layers in between the input and output layers [83]. Each layer contains a number of mathematical functions, whose output is an input to any number of functions in the next layer. A neural network is a method in deep learning that utilizes this principle and is modeled after the human neuronal pathway. In a neural network, the mathematical functions in intermediate layers are known as neurons or nodes. Intermediate layers beyond the first and last layer are known as hidden layers. The nodes are interconnected by the output of a previous node being processed as an input to the next node. This connection is known as an edge. Figure 2.5 shows an example of a neural network architecture.

The strength of a neural network is the complexity of the nodes specifically when they are interconnected. In general, this allows input data to be processed by different neuronal functions and the output to be combined with other neuronal functions. The large set of combinations results in the capability to correlate multiple features associated with a label. For instance, with a placenta ultrasound image, the neural functions may be a function to detect thousands of edge shapes. A neural network may be trained to select only the combination of edge shapes that are relevant to the placenta and use this combination to predict where the placenta is
Figure 2.5: An example of a neural network architecture. The circle represent nodes and lines represent the edges, which data from a previous node is processed as an input the the next node. The input layer is the column of nodes to the left and the output layer is the single column of node to the right. The edges are commonly assigned weights, which modify the input to the successor node. These weights are adjusted with each epoch until an optimized collection of weights configures the neural network to process input data to produce an output with optimized accuracy. The strength of a neural network is the interconnections in the hidden layers, which combine detailed feature computations for complex correlation of features to the output. Note that a neural network does not necessarily have only one input and one output node and can be an array of data.
on an image.

In neural networks, each neuron contains an activation function and the edge connecting to a successor neuron contains a weight, which is set to some initial value before training. The activation function of the successor neuron receives an input from the previous neuron that is modified by the weight. The modification can be a simple scalar multiplication or other regime in more complex weighting methods. The activation function uses the weighted input to compute an activated output. The activated output is then processed by the next edges and successor neurons. The selection of activation function varies depending on the application of the model and is an active area of research [47]. An intuitive activation function that simulates human neuronal pathways is the logistic, or sigmoid, or "soft step" function:

\[ \sigma(x) = \frac{1}{1 + e^{-(x-\theta)}} \]  

(2.1)

In the sigmoid function, the neuron is considered activated when the weighted input \( x \) is greater than a threshold value \( \theta \), where the unweighted output is the same as the weighted input. This mimics human neuronal pathways, where a signal must reach an electrical potential threshold before being propagated to the successor neuron.

### 2.5.3 Optimization of Neural Networks

The format of the output layer of a neural network varies depending on its application. The output layer may be a single neuron with an image of an outlined feature, or may be multiple neurons, each with a probability value of the input data being in a certain classification. In most cases, the optimization of the neuronal weights follow a similar regime. The output is compared to the ground truth and an optimization function determines the direction which to adjust the weights to maximize the accuracy.

The optimization is commonly performed by minimizing a cost function. The cost function is some metric that computes the inaccuracy of the predicted output. A common loss function, which is used in Chapter 5, is the binary cross entropy function:
\[ L = - \sum_{i=1}^{M} y_{x,i} \log(p_{x,i}) \]  

(2.2)

where \( M \) is the number of different labels, \( y \) is a binary value indicating if label \( i \) is the correct label for data \( x \), and \( p \) is the predicted probability that data \( x \) is labeled by label \( x \). This function is suitable for binary classification as it provides a logistic computation of the loss due to incorrect binary predictions.

This cost function is commonly used in segmentation of features or classification of features in images. In the case of segmentation, the pixels in an image serve as the input to a model and the model predicts if the pixel is within a segmentation or outside a segmentation. In the case of classification, the model predicts an entire image or patch of the image as belonging to a certain class.

To minimize the cost function, a family of techniques called backpropagation is employed [81]. Backpropagation relies on the principle that the input to a neuron \( N_i \) is the sum of the weighted outputs of all previous neurons it shared an edge with. By Gaussian error propagation, the partial error of \( N_i \) is then the linear combination of the partial error of previous neurons with respect to the weights. From the output neuron, we can iteratively compute the cost of the entire network with the weights as parameters. We can then employ techniques to optimize the weights and reduce the prediction error.

With \( n \) weights \( w_1, w_2, ... w_i \), the cost \( C(w_1, w_2, ... w_n) \) exists as a curve in \( n \)-dimensional space, containing several local minima and a global minima. Optimization functions can then be seen as a geometric problem to find the global minimum of a curve. One of the earliest and basic optimization functions to illustrate optimization is gradient descent [7]:

\[
(w_{1,i+1}, w_{2,i+1}, ... w_{n,i+1}) = (w_{1,i}, w_{2,i}, ... w_{n,i}) - \alpha \nabla C(w_{1,i}, w_{2,i}, ... w_{n,i})
\]

(2.3)

where \((w_{1,i}, w_{2,i}, ... w_{n,i})\) is the set of weights in iteration \( i \), \( C(w_{1,i}, w_{2,i}, ... w_{n,i}) \) is the cost function, and \( \alpha \) is a step size parameter modifying the impact of the gradient of the cost function, known as the learning rate. The gradient descent function is iterative, and continues until a maximum iteration cycle count or if the weights do
not change by more than a threshold value.

The intuition of gradient descent is that for a multivariate function $C(w_{1,i}, w_{2,i}, ..., w_{n,i})$ that is differentiable at point $(w_{1,i}, w_{2,i}, ..., w_{n,i})$, the function decreases the fastest in the direction of $-\nabla C(w_{1,i}, w_{2,i}, ..., w_{n,i})$. The learning rate aims to adjust the magnitude of the step sizes in descending the multivariate curve, as too large of a step size results in overshooting a minimum and too small of a step size results in increased computational time. Although this method is simple, it is susceptible to errors such as converging in a local minimum instead of the global minimum.

The selection of the optimization function and learning rate is important, different methods have advantages depending on the application. Chapter 4 discusses the selection of specific optimization functions for tasks of segmentation and classification.

### 2.5.4 Convolutional Neural Networks

Convolutional neural networks CNN are a class of neural networks where the activation output of a neuron is passed through a convolutional kernel in a convolutional layer before being weighted and passed to the successor neuron, as illustrated in Figure 2.6. CNNs are typically used for higher dimensional data, such as for images. The input to a CNN is an $n$-dimensional tensor, which undergoes a convolution with an $n$-dimensional convolutional kernel tensor. This is advantageous as the convolutional kernels can be selected to be associated with specific features such that each neuron can be interpreted as the kernel which identifies if the feature is important in predicting the output label. For instance, in an image classification CNN, for classification of lesions placenta ultrasound, one kernel may be a matrix that, when convolved with an image, filters out all regions that are homogeneous within a certain size. If the optimization of this network results in the kernel being weighted significantly, then this can be interpreted as prediction of lesions relying on identifying regions that are inhomogeneous.

CNNs can be designed with a variety of layers, convolutional kernels, activation functions and network parameter such as the learning rate. CNNs are much slower than lower-dimensional neural networks due to the complexity of the convolutional layers and input data. However, with recent advances in computational power,
A visualization of a CNN. A CNN is different from lower dimensional neural networks in that the input is more complex, usually existing as a higher dimensional tensor. The data is processed through neurons with convolutional kernels, which are similar dimensional tensors that convolve with the data from a previous neuron. The advantage of this process is that the output of a convolutional layer represents highly complex feature data, which can be used for detailed correlation to predict a label. A CNN usually contains other utility functions, such as max-pooling to decrease the size of images or densification to combine the width of layers.

particularly for graphics processing unit (GPU)s [69], CNNs have recently shown success for real-time performance of several tasks in medical imaging. CNNs have been successful in segmentation [82] and classification [6] for complex medical images due to the ability to extra highly detailed features through its convolutional kernels. Design of an accurate CNN is important for its utilization in ML techniques for placental ultrasound. Specific applications of CNNs are discussed in Chapter 4 and Chapter 5.
Chapter 3

Automated Acoustic Shadow Detection

3.1 Introduction

The unique artifacts in ultrasound increases the difficulty in interpreting ultrasound images, particularly for training machine learning algorithms, where a large set of data is required and artifacts may obscure expected anatomy. An acoustic shadow is one artifact that is particularly troublesome as it results in a region of signal loss. This section discusses the background of acoustic shadows and presents two methods to detect acoustic shadows automatically. The presented methods were designed to be easily usable without parameter tuning for each image such that placenta ultrasound images can be automatically processed to highlight ultrasound artifacts. The highlighted artifacts can then serve as an input for further ML methods, which ultimately provide quantified placenta analysis to aid physicians in assessing the placenta.

3.1.1 Background of Acoustic Shadows

An acoustic shadow occurs when an acoustic wave from a transducer surface propagates towards a boundary of two materials with significantly different wave impedance properties [51]. The ultrasound wave is then mostly reflected, resulting in a loss of
Figure 3.1: Example of an acoustic shadow in an arm ultrasound. An acoustic shadow is created at bone-tissue boundary of the radial joint, radius, and ulna, resulting is a bright region followed by an almost total signal loss further down the scanline. As a result, anatomy beyond the shadow boundary is obscured.

Acoustic shadows are significant as they are an asset and detriment in ultrasound imaging. Acoustic shadows are known to occur at air-tissue, tissue-bone, and tissue-lesion interfaces [34]. By observing an acoustic shadow, a user can identify the presence of an air gap due to incomplete transducer contact, calcifications in tissue, or lesions depending on the context of the imaging location. Acoustic shadows have been leveraged to identify gall stones [34] or track the size of kidney stones [29].

The interpretation of acoustic shadows is commonly performed by manual inspection from an expert, such as a radiologist. However, there are several motivations for the automatic detection of acoustic shadows. First, users with less sonog-
raphy experience could interpret ultrasound imagery easier with acoustic shadows automatically identified. Secondly, acoustic shadows limit the capability of automated image processing. Modern analysis techniques involving images with shadows such as 3D reconstruction [67], fiducial registration [101], and training of supervised learning algorithms [33] require features to be identified or flagged as a shadow, which is time consuming if done manually.

3.2 Related Work

Several methods have been used in literature to detect shadows and illustrative examples are discussed. Geometric techniques model the path of an ultrasound signal for an expected image along the scanline using a random walk [46]. Pixels are then flagged as a shadow if it is below a heuristic confidence threshold of 0.25. However, geometric techniques require knowledge of ultrasound transducer properties to parameterize random walk weights, such as the focal length, radius of curvature, and thickness. The technique is therefore challenging to implement across different ultrasound equipment, particularly for applications such as fetal imaging, which is performed in a variety of clinics. This also reduces applicability for machine learning applications as accurate transducer parameter labels are required for each image.

Pixel gray level methods ignore the transducer properties and analyze only the graphical properties of an image [38]. Shadows have been detected on brain images by analyzing the entropy along a scanline to flag pixels of sudden low entropy as a potential shadow. These techniques were validated with the Dice coefficient, which computes the similarity of the predicted shadow regions and the manually segmented shadow regions. Hellier et al. [38] achieved a comparable Dice similarity coefficient as geometric methods but required specific thresholding, window sizing, filtering, and image mask parameterization for different anatomy and transducers. The drawback was again the need for parameterization and tuning, which requires image processing expertise and prior knowledge of specific applications.

ML methods have gained significant interest in medical imaging analysis. To our knowledge, no machine learning method has demonstrated the capability of general shadow detection from multiple types of anatomy. Deep learning meth-
ods have identified features in a specific image sets that contain shadows, such as neuroanatomical regions in cranial scan [62] or spinal levels in a posterior scan [39]. Although ML has the potential of providing automated feature recognition in multiple applications, a ML algorithm benefits the most from a large dataset. Ultrasound imaging is highly variable due to unique artifacts, operator techniques, and equipment. In addition, shadows are a common feature that occur in various imaging scenarios. Previous techniques focused on a single anatomical region and training data was from a consistent imaging scenario. However, it is difficult to construct a training data set with the generality required to recognize shadows in different scenarios usable for a variety of ultrasound applications.

There are two objectives to this chapter. First, to address the need for understanding general characteristics of shadows, a study was conducted to scan multiple anatomy and transducers specifically to analyze the statistics of different types of shadows. Second, to address existing needs for versatile detection with minimal parameterization, previous methods were then extended utilizing statistical thresholding of RF or B-MODE data to detect shadows from various imaging scenarios. The two methods are illustrated in a flowchart in Figure 3.2.

3.3 Materials and Methods

3.3.1 Data Collection

Ultrasound B-MODE and RF data were acquired by scanning 37 adult participants with informed written consent under a clinical study approved by the University of British Columbia Clinical Research Ethics Board (ID: H18-00368). Each patient was scan on the forearm near the distal end of the pronator quadratus in the supinated position, on the elbow near the cubital fossa in the supinated position, and on the anterior surface of right ribs 11-12 in a laid down position. The scanning locations are visualized in Figure 3.3.

The scans were taken with a curvilinear (C5-2/60, Ultrasonix, Richmond, British Columbia, Canada) and a linear (L14-5/38, Ultrasonix, Richmond, British Columbia Canada) transducer for a total of 6 images per participant. Different transducer settings were used for each anatomical region and transducer, summarized in Table 3.3.
**Figure 3.2:** Processing steps for RF and B-MODE shadow detection. RF processing is used if RF data is available and involves fitting the Nakagami distribution onto the echo envelope of each RF scanline before adaptive thresholding with Otsu’s method. In many cases, there may only be access to B-MODE image data, for which an entropy map is computed and similar adaptive thresholding is used to detect shadows.

**Table 3.1:** Transducer properties for different imaging scenarios.

<table>
<thead>
<tr>
<th>Anatomy</th>
<th>Frequency</th>
<th>Depth</th>
<th>Gain</th>
<th>Pulse Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linear Transducer (L14-5/38)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forearm</td>
<td>11.0MHz</td>
<td>5.0cm</td>
<td>50%</td>
<td>0.6mm</td>
</tr>
<tr>
<td>Elbow</td>
<td>11.0MHz</td>
<td>5.0cm</td>
<td>40%</td>
<td>0.6mm</td>
</tr>
<tr>
<td>Ribcage</td>
<td>5.0MHz</td>
<td>10.0cm</td>
<td>30%</td>
<td>1.7mm</td>
</tr>
<tr>
<td><strong>Curvilinear Transducer (C5-2/60)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forearm</td>
<td>4.0MHz</td>
<td>5.0cm</td>
<td>50%</td>
<td>2.6mm</td>
</tr>
<tr>
<td>Elbow</td>
<td>4.0MHz</td>
<td>5.0cm</td>
<td>40%</td>
<td>2.6mm</td>
</tr>
<tr>
<td>Ribcage</td>
<td>3.3MHz</td>
<td>10.0cm</td>
<td>30%</td>
<td>5.5mm</td>
</tr>
</tbody>
</table>

**Table 3.1**

Shadows were expected to occur due to superficial and deep bones and from an air gap created by the lateral edges of the transducer not being in flush contact with the skin. The experiment was designed to generate a dataset from various imaging scenarios to explore general shadow characteristics and to validate the versatility...
Figure 3.3: Visualization of the three scanning locations in the clinical study including the a) forearm, b) elbow, and c) ribcage with corresponding images in d), e), and f). Acoustic shadows are expected due to the bone-tissue boundary in the arm and ribs.

3.3.2 RF Speckle Detection

To analyze shadows, windows of speckle were analyzed on the RF signal. Speckle occurs from interference of randomly distributed microscopic scatterers, resulting in a granular appearance on the image [13]. To produce B-MODE images, manufacturers often employ image enhancement algorithms, such as logarithmic com-
Figure 3.4: A comparison of a scan of the forearm from a linear transducer with the a) brightness-mode and b) logarithmic echo envelope of the RF data. The B-MODE data is visually enhanced to remove the speckle for a smoother image, but the speckle in the RF data can be leveraged to understand the scattering interactions of the acoustic wave in certain materials.

pression, nonlinearly alter speckle patterns. B-MODE image formation can also be manipulated by an operator to visually enhance an image, such as adjusting time-gain compensation or dynamic range. Thus, the underlying speckle analysis in RF signals can provide shadow detection usable across different machines and operators. However, the original speckle pattern contains information related to the acoustic interactions in tissue [13]. By analyzing the RF signal distribution, we can statistically characterize the distributions in tissue compared to shadow regions. A comparison of B-MODE and RF images is shown in Figure 3.4.

We expect tissue to resemble speckle modeled by known distributions and expect shadow to resemble different distributions, which may be a mixture of lessened speckle due to the signal loss and background electronic noise. Previous
studies have attempted despeckling methods on images containing shadows [5] by using filters based on a Rayleigh-like distribution. As such, even if shadow regions do not exactly resemble known speckle distributions, they may still be characterized to a sufficient extent with known distributions for a maximum likelihood fit. The fitted parameters can then be used to differentiate between shadow and non-shadow regions.

To understand the behaviour of speckle, we can model speckle as the sum of scatterers within a cell size. The scatters are considered a random walk as the backscattering of acoustic waves is considered to be random. We can then expressed the scattered acoustic wave as in Equation 3.1

\[
\sigma(t) = \sum_{n=1}^{N} A_n \cos(\omega_0 t + \phi_n)
\] (3.1)

where \(\omega_0\) is the mean frequency, \(N\) is the number of scatters within the cell, \(A_n\) is the amplitude, and \(\phi_n\) is the phase of scatterer \(n\). At large \(N\), we can utilize Ptolemy’s identities to rewrite this as Equation 3.1

\[
\sigma(t) = \sum_{n=1}^{N} A_n (\cos(\omega_0 t)\cos(\phi_n) - \sin(\omega_0 t)\sin(\phi_n))
\] (3.2)

\[
\sigma(t) = X \cos(\omega_0 t) - Y \sin(\omega_0 t)
\] (3.3)

where \(X\) and \(Y\) are coefficients. With large \(N\) and due to the periodicity of the sinusoidal terms, \(X\) and \(Y\) are expected to have zero mean with a Gaussian distribution due to the law of large numbers.

\[
X = \sum_{n=1}^{N} A_n \cos(\phi_n)
\] (3.4)

\[
Y = \sum_{n=1}^{N} A_n \sin(\phi_n)
\] (3.5)

We can then compute the envelope of the backscattered signal in Equation 3.6

\[
S(t) = \sqrt{X^2 + Y^2}
\] (3.6)
One of the first models for the amplitude of this envelope is the one parameter Rayleigh distribution to model the probability density of a random walk with non-negative random values [13] in Equation 3.7

\[ p(A) = \frac{A}{\Phi} e^{-\frac{A^2}{2\Phi}} \quad (3.7) \]

where \( A \) is the RF intensity and \( \Phi \) is the scale parameter.

The Rayleigh distribution is capable of modeling fully developed speckle, which does not occur in limited scattering [95]. More generalized models have been applied such as the Rician, Homodyned-K, and Nakagami distributions to characterize speckle [28]. The utility of speckle has been demonstrated in the literature to classify tumorigenicity of breast lesions [16] or levels of liver fibrosis [40] by categorizing image regions based on the speckle pattern. Shadow characterization presents a simpler problem as shadow and non-shadow regions contain significantly different speckle patterns. Thus, the Nakagami distribution expressed in Equation 3.8 was chosen to model speckle. The Nakagami distribution provides greater generality than the Rayleigh distribution while being more computationally efficient than the Rician or Homodyned K distributions [28] and is shown in Equation 3.8

\[ \Phi(x, m, \omega) = 2\left(\frac{m}{\omega}\right)^m \frac{1}{\Gamma(m)} x^{2m-1} e^{-\frac{m}{\omega}x^2} \quad (3.8) \]

where \( x \) is RF intensity, \( m \) is the shape parameter or Nakagami \( m \) parameter, \( \omega \) is a scale parameter and \( \Gamma(m) \) is the gamma distribution.

To characterize shadows, the raw RF data was first processed by computing the echo envelope of each scanline with a Hilbert transform by Equation 3.9. The processing is visualized in Figure 3.5

\[ H(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(x)}{t-x} dx \quad (3.9) \]

This was performed on an averaged RF signal from three image frames. This creates a pre-scan converted image, visually similar to B-MODE but without filtering to alter speckle. Next, the RF image was divided into overlapped windows with a width of a single RF scanline and a length of three times the pulse length.
Figure 3.5: The processing method for RF data with a) the raw RF data with a sample scanline highlighted, b) the echo envelope of the raw RF data scanline, c) the echo envelope of the entire RF image, and d) the logarithmic scaled echo envelope image. This processing is required for the RF data to contain statistically distinguishable characteristics between patches to characterize acoustic shadows.

We expect the width of a single RF scanline to be on the order of magnitude of a resolution cell, which is on the same order of magnitude as the correlation length [97]. The window length was demonstrated in literature to be sufficiently large to capture multiple wavelengths and scattering events while being small enough to be useful in differentiating different regions on the millimeter scale [16]. The echo envelope processing is shown in Figure 3.5.

Next, each window was fit to a Nakagami distribution using a maximum likelihood estimate to compute a map of Nakagami parameters $m$ and $\omega$, as shown in...
Figure 3.6: A visualization of the B-mode and RF parameter maps. The b) Entropy Map was computed from processing of the a) original B-MODE image and the d) Nakagami $\omega$ map was computed from the c) echo envelope. Note that the echo envelope contains noticeable speckle, which has been used to fit a Nakagami distribution to characterize shadow. The region at depth 2.50 cm and scanlines 32-40 is attenuation and not a shadow. This is an important distinction in shadow detection and both maps show the region as below a threshold to flag a shadow boundary.

Then, for each ultrasound image, Otsu’s method was applied to its Nakagami $\omega$ map to automatically compute a $\omega$ threshold for each individual image as we expect separate distributions for shadow and non-shadow regions. Otsu’s method is expressed in Equation 3.10

$$\tau|_{\omega_i^2(t)} = \min(\omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t))$$

(3.10)
\[ \omega_0(t) = \sum_{i=0}^{t-1} p(i) \]  

\[ \omega_1(t) = \sum_{i=0}^{L-1} p(i) \]  

(3.11)  

(3.12)

where \( \tau \) is the computed threshold, \( t \) are the iterated values of Nakagami parameters to search for a threshold, and \( L \) is the number of bins in the distribution of Nakagami parameters.

This was sufficient as the \( \omega \) parameter is significantly different for shadow regions with abundant speckle and non-shadow regions with minimal speckle, Then, for each scanline, the axially deepest data point that is above the threshold is labeled as the shadow boundary and all data points below are labeled as a shadow.

The Nakagami shape parameter, \( m \), was also investigated, though there was not sufficient delineation between parameter values in shadow and non-shadow regions for this parameter to be effective in thresholding. The distributions of the two parameters are displayed for shadow and non-shadow regions in Figure 3.10.

### 3.3.3 B-MODE Entropy Detection

Many ultrasound machines do not provide access to RF data for speckle analysis. Thus, a previous pixel gray level shadow detection method on B-MODE images was modified and extended. Scanline entropy was investigated on B-MODE images to characterize different types of shadows, but with the addition of adaptive thresholding of entropy to address the need for usability with minimum configuration. B-MODE analysis was performed on an averaged image from three image frames, similar to RF analysis. First, the cumulative scanline entropy is computed for each pixel, similar to the “Rupture Criterion” [38], with approximate window size fixed as the size of a resolution cell similar to the RF analysis method. The entropy computation is shown in Equation 3.13

\[ S_{i,j} = \sum_{k=1}^{3n} \frac{I(i-k,j) \log_2 \frac{I(i-k,j)}{I(i+k,j)}}{I(i+k,j)} + \frac{I(i+k,j) \log_2 \frac{I(i+k,j)}{I(i-k,j)}}{I(i-k,j)} \]  

(3.13)
where $S_{i,j}$ is the cumulative entropy at pixel $i$ on scanline $j$, $\eta$ is the pulse length, and $I(i)$ is the gray level (0-255) of pixel $(i,j)$. For the case of curvilinear images, radial scanlines were linearly interpolated between the two symmetric lateral edges of the image.

For images from the linear transducer, identifying the path of the scanline is simple as it corresponds to the vertical columns of the image. For the curvilinear transducer, however, the scanlines need to be estimated for the image. It is possible to deduce the geometric transformation from the linear RF data to the curved B-MODE image by creating a map between the data points. However, this is cumbersome and decreases the usability of this method as new maps are required for different transducers. Instead, a simple linear extrapolation is performed to estimate the direction of the scanline.

For curvilinear images, the algorithm first looks for all the pixels which represent the beginning of a scanline. From the upper left corner, the algorithm then tracks the ultrasonic ring-down, which is a persistent artefact that produces bright bands for a small thickness starting at the top of the image caused by fluid in the transducer that emits a continuous signal to the transducer [3]. With the slope identified on the lateral edge and assuming that the transducer produces a symmetrical image, the algorithm creates scanline paths originating from the first imaging pixel at the top of the image. The slope of the paths are linearly interpolated between the two slopes of the two lateral edges. Scanline tracking is visualized in Figure 3.7.

Next, Otsu’s method is applied onto the entropy map of each image to automatically compute a threshold entropy value, similar to RF analysis. The intuition of the threshold is different than in RF analysis. In RF analysis, the threshold separates patches of intense and minimal speckle. In B-mode analysis, the threshold separates pixels of a shadow boundary, which has high entropy, and pixels away from shadow boundary, which include shadow and non-shadow regions. Thus, shadows can be identified by finding the last pixel on a scanline with an entropy higher than the threshold, representing a bright shadow boundary.
Figure 3.7: Creating the scanline paths on a curvilinear ultrasound image of the right ribcage and the forearm with different depth settings on the curvilinear transducers. Two initial paths are created by computing the slopes of the two lateral edges, which are detected by tracking the ringdown artifact. The remaining paths are created by linearly interpolating the slope of pixels in between the lateral edges.

3.3.4 Validation

A trained annotator manually outlined the boundary of the shadow regions on B-mode images. The manual regions were used as a gold standard, as manual identification is common in clinical practice and has been used in previous literature for comparison [38]. A Dice coefficient was computed to compare similarity of manual and automated shadow detection. The manual outline was used to define four regions for classification of statistical parameters: a non-shadow region above the boundary, a shadow region below the boundary, a “transition region”, which is a window defined as three resolution cells long axially below the boundary, and a “deep shadow region”, which is the data below the transition region. The validation was repeated with the RF and entropy window increased and decreased by 50%. The Ljung-Box Q-test was used to measure residual autocorrelation of the Dice
Table 3.2: Mean Dice coefficients for different imaging scenarios ± standard deviation.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>RF</th>
<th>B-Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear (L14-5/38)</td>
<td>Forearm</td>
<td>0.91±0.05</td>
</tr>
<tr>
<td></td>
<td>Elbow</td>
<td>0.94±0.06</td>
</tr>
<tr>
<td></td>
<td>Ribcage</td>
<td>0.87±0.09</td>
</tr>
<tr>
<td>Curvilinear (C5-2/60)</td>
<td>Forearm</td>
<td>0.89±0.05</td>
</tr>
<tr>
<td></td>
<td>Elbow</td>
<td>0.93±0.04</td>
</tr>
<tr>
<td></td>
<td>Ribcage</td>
<td>0.83±0.08</td>
</tr>
<tr>
<td>Mean All Anatomy</td>
<td>All Anatomy</td>
<td>0.90±0.07</td>
</tr>
</tbody>
</table>

coefficients. A Wilcoxon rank sum test has been performed between Nakagami ω parameter values in shadow and non-shadow regions and between entropy values in shadow and non-shadow regions.

As an initial experiment, a gelatin phantom was created with slits of wooded embedded at 0.75cm and at 2.50cm to create a region of shallow and deeper shadows on both edges of the phantom. The gain was varied and both RF and B-MODE methods were employed to test the feasibility of the methods on a clearly visible shadow, shown in Figure 3.8. When comparing to manual segmentation, all detected shadows resulted in a Dice coefficient of above 0.95, with the lowest score being the entropy method applied on a high-gain image. This provides support that extreme operator adjustments on the B-mode image may affect pixel gray level detection methods more than RF methods.

3.4 Results

The detection example from the initial gelatin phantom experiment is visualization is Figure 3.8. Examples of detected shadows from both methods are highlighted in gray in Figure 3.9 in different shadow detection scenarios. The Dice coefficients for both methods for different anatomy and transducers are shown in Table 3.2.

The mean Dice coefficients (± standard deviation) were 0.90±0.07 and 0.87±0.08 for RF and B-mode methods. Manual annotation was repeated five times with a mean Dice coefficient of 0.92±0.02 for all images and transducers. The Dice coefficient did not change by more than 0.03 when the window size was varied by
**Figure 3.8:** Images of both RF and B-MODE shadow detection performed on a gelatin phantom with two wooden slits embedded at a depth of 0.75cm and 2.50cm. The phantom was made to simulate shallow, deep, and non-shadow regions. The methods were capable of shadow detection with a Dice coefficient > 0.95, though noticeable errors were present at high-gain images for the B-mode method. This is expected as B-MODE methods rely on pixel gray level, which may vary due to operator settings.

With the benefit of a varied dataset, general statistics of shadows can be analyzed, as summarized in Table 3.3 and Table 3.4. The distributions of Nakagami parameters and entropy for the different regions are visualized in Figure 3.10.

For shadow detection, the parameters differentiating a shadow and non-shadow are of particular interest. Shadows were observed to have a mean Nakagami $\omega$ parameter of $4.14 \pm 0.40$ and a mean entropy of $1.03 \pm 0.29$ whereas non-shadows were observed to have a mean $\omega$ of $6.24 \pm 0.92$ and $2.20 \pm 0.81$. Wilcoxon rank sum $p$ values were less than 0.002 between Nakagami $\omega$ parameter distributions in shadow and non-shadow regions and less than 0.001 between entropy distribu-
Figure 3.9: A comparison of the original B-mode images, the detected shadows manual detection, RF detection, and B-mode detection. Both detection methods perform similarly to manual detection. Both methods perform slightly less accurately on curvilinear images, potentially due to the reduced resolution from interpolating the scanlines. Most errors of RF detection occur near the shadow boundary, potentially due to the transitioning speckle from non-shadow to shadow.

3.5 Discussion

The RF and B-MODE shadow detection developed achieved a comparable Dice similarity coefficient to manual detection for all anatomy and transducer types.
Table 3.3: The mean Nakagami $\omega$ and Entropy values of different anatomy, transducer, and shadowing region ± standard deviation. Values are consistent among different transducers and anatomical regions. The variance of entropy and Nakagami $\omega$ in one imaging region and transducer setting is less than the variance across different regions and transducers for shadows and non-shadows.

<table>
<thead>
<tr>
<th></th>
<th>Linear (L14-5/38)</th>
<th>Curvilinear (C5-2/60)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forearm</td>
<td>Elbow</td>
</tr>
<tr>
<td>Nakagami $\omega$ (Log Scale)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shadow</td>
<td>4.15 ± 0.45</td>
<td>4.18 ± 0.45</td>
</tr>
<tr>
<td>Non-Shadow</td>
<td>6.19 ± 0.96</td>
<td>6.49 ± 0.97</td>
</tr>
<tr>
<td>Transition</td>
<td>4.94 ± 0.62</td>
<td>5.36 ± 0.62</td>
</tr>
<tr>
<td>Deep Shadow</td>
<td>4.13 ± 0.43</td>
<td>4.16 ± 0.43</td>
</tr>
<tr>
<td>Entropy (Log Scale)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shadow</td>
<td>0.92 ± 0.22</td>
<td>1.10 ± 0.36</td>
</tr>
<tr>
<td>Non-Shadow</td>
<td>2.34 ± 0.96</td>
<td>2.34 ± 0.80</td>
</tr>
<tr>
<td>Transition</td>
<td>2.45 ± 0.62</td>
<td>2.56 ± 0.53</td>
</tr>
<tr>
<td>Deep Shadow</td>
<td>0.71 ± 0.43</td>
<td>0.89 ± 0.26</td>
</tr>
</tbody>
</table>

Table 3.4: The mean Nakagami $\omega$ and Entropy values of all anatomy and transducers for different shadowing regions ± standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Mean Nakagami $\omega$ (Log Scale)</th>
<th>Mean Entropy (Log Scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shadow</td>
<td>4.14 ± 0.40</td>
<td>1.03 ± 0.29</td>
</tr>
<tr>
<td>Non-Shadow</td>
<td>6.24 ± 0.92</td>
<td>2.02 ± 0.81</td>
</tr>
<tr>
<td>Transition</td>
<td>5.08 ± 0.77</td>
<td>2.21 ± 0.84</td>
</tr>
<tr>
<td>Deep Shadow</td>
<td>4.06 ± 0.34</td>
<td>0.89 ± 0.27</td>
</tr>
</tbody>
</table>

(p < 0.025). The previous studies using B-MODE entropy reported a mean Dice coefficient of 0.91±0.07 between manual annotators [38]. An important feature of shadow detection is being able to differentiate between a shadow and simply high attenuation of the signal. Both scenarios result in an eventual loss of signal. Shadow detection, however, has a characteristic high gray level shadow boundary before a significant loss in signal, compared to gradual signal losses in attenuation. The high Dice similarity coefficient indicates that both methods were capable
Figure 3.10: Histograms of Nakagami parameters and entropy values in shadow and non-shadowing regions. The Nakagami $\omega$ and Entropy distributions have a more noticeable delineation between shadowing and non-shadowing distributions compared to the Nakagami $m$ parameter, which was not used to threshold shadow boundaries. Entropy is very minimal in continuous dark shadow regions, which is expected due to the minimal variations in pixel gray level.

of this distinction. This is also visualized in Figure 3.9, where regions of low gray level without a bright shadow boundary were correctly labeled as non-shadow. The accuracy supports the versatility of the detection method as both methods are able to identify shadows across different anatomy and transducers with minimum configuration.

For a general observation for shadows, the computed Nakagami $\omega$ parameters of all manually outlined shadows indicate that there is a statistically significant difference between shadow and non-shadow regions, regardless of anatomy and transducer and even with the error in the transition regions considered. The speckle and its statistics from shadows is thus distinct from the speckle created by tissue, muscle, or fat. This observation can be utilized in the future for further analysis of shadows.
In RF detection, both false positive and false negative errors most frequently occurred immediately below a shadow boundary as opposed to B-mode detection where errors were in various regions. To study the frequent areas of error further, the “transition region” immediately below a manually annotated shadow boundary and a “deep shadow region” below the transition region was investigated. The Nakagami $\omega$ parameter of transition regions of all anatomy and transducers were within a standard deviation of both shadow and non-shadow regions. The deeper shadow regions were observed to have a lower Nakagami $\omega$ parameter than shadow regions and with a lower standard deviation. The spread of the speckle also significantly decreases after. This indicates that the transition region cannot be fully distinguished from either a shadow or non-shadow and presents as it is statistically similar to the two. This is potentially the cause of the errors, as the speckle distribution is much more consistent in the deep shadow regions compared to any other region. Physically, speckle interactions appear to gradually lessen after a brightest point on a scanline, possibly due to incomplete total reflection at a boundary. The boundary is thus not an instantaneous division between non-shadow and shadow, rather, there is a transition region with statistics between a shadow and non-shadow before the speckle fully resembles a shadow.

In the transition region of B-MODE images, the entropy values were similar but consistently higher than non-shadow values. This is expected as entropy is the highest when there is the greatest change in pixel gray level, which occurs at a shadow boundary, even with the a non-instantaneous non-shadow to shadow transition. However, the averaged entropy of all non-shadow regions have a greater spread than the Nakagami parameters, likely due to the differing operator settings used. Thus, B-MODE detection may not be as consistent as RF detection.

3.6 Limitations

In our study, although a range of frequencies and equipment were used, the parameters were still limited and not all combinations were explored. To further validate the detection method, future work would include a more extensive investigation of these parameters, such as with a random parameter grid search, to provide more support for widespread clinical use.
As both RF and B-MODE images search for a threshold for the shadow boundary, it is possible to misinterpret a reverberation artifact as a beginning of a shadow. Reverberation at a shadow boundary would cause a similar bright region followed by a dark region, which visually appears like a shadow boundary despite being an artifact in a shadow region. This is a limitation in our method and future work includes integration of reverberation identification, such as identifying echo time duration to know what pulses correspond to anatomical interaction [99], would be required to reduce reverberation errors.

There is a limitation with analysis using the Nakagami distribution in that the fitted Nakagami distribution to model scatterers change depending on transducer frequency. Previous literature observed that in the 36-58MHz frequency range, the Nakagami $m$ parameter decreased near the theoretical lower limit compared to a higher Nakagami $m$ parameter value at 10MHz signal [23]. This was reported to be due to the spatial organization of the cells being ”on the order of a fraction of the wavelength” and a Nakagami distribution cannot model the scatterers of red blood cells at this frequency. Due to this and from limitations of the equipment used in our study, we cannot conclude that shadow detection with Nakagami analysis will be accurate in higher frequencies beyond the values tested. Future studies are required to analysis the performance of shadow detection in higher frequencies. Diagnostic ultrasound commonly uses a frequency range of 2-15MHz [45] and higher frequencies are limited to subspecialized cases such as optical ultrasound [71]. Shadow detection is expected be applicable in most use cases without issues from the high frequency behaviour of the Nakagami distribution.

There is a limitation for diagnostic usage of the proposed shadow method in cases where acoustic shadowing does not exhibit the characteristic bright boundary followed by a dark region. In cases where there is partial or incomplete shadowing, such as small calcifications in the placenta [1]. In these cases, there is a resemblance of a shadow, where the calcification is brighter and the region below is noticeably darker, but not with a brightness difference as extreme as shadowing from the ulna and the regions below retain speckle similar to tissue. Although calcifications are pathologically important to recognize, the proposed shadow detection method would likely be unable to detect the partial shadowing from these calcifications. The proposed method would be applicable only in cases of more
complete shadowing, which would still be practical for significant gall and kidney stones, for instance.

In previous literature, shadows were defined qualitatively [51] as a sudden loss of signal and brightness. The observed transition region in this study suggests that the qualitative definition of a shadow may be insufficient for accurate detection. One algorithm may detect the shadow starting immediately after the brightest location, or another may use a convention such as a full width at half maximum to define where the signal has sufficiently low gray level to resemble the start of a shadow. There is a decision point required for a clear definition for where a shadow begins to improve shadow detection accuracy, both from a signaling perspective for image processing and a visual perspective for manual inspection.

The findings in this study result in several implications. First, the statistics of acoustic shadows have been investigated on a dataset with shadows occurring from multiple scenarios as opposed to specific cases where shadows are observed. This provided a more generalizable observation that shadows can be characterized by distinctive speckle distributions in different of anatomy and equipment and that there exists a transition region before the loss of speckle in a shadow. Second, the shadow detection methods demonstrated accuracy similar to previous methods [38], indicating that the same shadow detection method can be used with different transducer or imaging location. In future studies, the speckle statistics observed can be used to develop further models for anatomical features containing shadows. In ML algorithms, an initial network could be used with the shadow detection methods presented. Future studies would also have to take into consideration the most frequent source of error of shadow detection as the shadow boundary.

3.7 Conclusion

Acoustic shadows from different imaging scenarios were investigated. RF and B-MODE methods were developed for acoustic shadow detection requiring only the transducer pulse length as the input parameter. When comparing to manual detection, the methods achieved a Dice similarity coefficient within range of manual observers. The work focused on applying shadow detection and statistical analysis to a varied dataset of three different anatomical locations and two different
transducer to provide a representative understanding of general acoustic shadows. The statistics of acoustic shadow indicate that shadows contain a distinct speckle distribution compared to non-shadows and the speckle characteristics transition at the shadow boundary. The statistical findings of shadows can aid interpretation of ultrasound images in the future using speckle analysis. The versatility of the shadow detection method has the potential to improve the interpretation of ultrasound images with shadow artifacts or to serve as a pre-processing step for ML methods.
Chapter 4

Automated Placenta Segmentation

4.1 Introduction

In developing ML methods for classifying a placenta as normal or abnormal, the preprocessing step of extracting the placenta is first required. This section discusses the motivation of placenta segmentation, the CNN developed, the results of a clinical study, and the limitations of the method.

4.1.1 Motivation for Automated Placenta Segmentation

Without extraction of the placenta, features not related to the placenta, such as adjacent fluid or fetal anatomy, may confound studies that attempt to correlate placenta features with clinical outcomes. In addition, processing an entire ultrasound image is more time consuming than processing only the placenta region. Placenta segmentation can be performed manually, but this poses several challenges.

First, extracting the placenta from fetal ultrasound images is time-consuming if performed manually. Applying ML to correlate placental images with disease requires a large dataset as ultrasound images vary in appearance depending on the type of machine, transducer settings, and operator technique. An automated algorithm for placental detection offers the advantage of rapid detection of complete
and unselected placental tissue for further analysis.

Second, the reliance on visual cues to detect the placenta may result in different segmentations by different observers. In some cases, borders and the interface between myometrium and adjacent fetal parts may be difficult to determine. Features such as a dark infarction on the edge may be interpreted as non-placental fluid. Early in the pregnancy, the placenta may be irregularly shaped and lack an obvious border to uterus [4], limiting the ability to study the placenta early.

Third, there is an increased number of novice ultrasound users from its decreasing cost, particularly in low-resource areas and for point-of-care ultrasound [17]. Without specialized training, placenta identification and analysis is challenging. Automated placental detection and assessment is particularly amenable to use in remote areas where there is no subspecialized expertise, possibly preventing unnecessary travel to specialized centres.

4.2 Related Work

There are several existing studies on automated placenta detection. Gupta et al. [36] proposed texture-based conditional random fields to deduce the placenta from the presence of surrounding maternal tissue, tested on two fetal images. Stevenson et al. presented a semi-automatic method for 3D placenta segmentation with accuracy similar to manual segmentation [91]. However, this required manual initialization to seed the random-walker algorithm. A fully automatic method was developed by Looney et al. [59] using a CNN to segment placental volume in 3D ultrasound images, obtaining a Dice coefficient of 0.73. However, this required down-sampling 3D volumes and training with 240 images due to limitations of hardware. Only one transducer and machine type was used to generate the dataset analyzed. We sought to build on these described methods and develop a fully automated solution for two-dimensional (2D) imaging with higher accuracy, validated on a diverse dataset.

There are two main contributions of the work from this chapter. First, we propose a detection method by modifying an existing CNN to be combined with a layer for acoustic shadow detection based on analysis of ultrasound physics. Second, we acquired a dataset of 1364 2D fetal ultrasound images from different opera-
tors, machines, and patients to develop the neural network. This dataset was used specifically to develop a method usable across different scenarios, as parameters such as the brightness, imaging placenta, and if the placenta is not fully in view may vary. Ultimately, automated placental segmentation is the first step towards the creation of a widely applicable, low cost, objective screening tool for future studies in the detection of placental pathology with artificial intelligence methods requiring large-scale data processing.

4.3 Materials and Methods

4.3.1 Data Collection

The acquisition of images was approved by the University of British Columbia Clinical Research Ethics Board (ID: H18-01199). 1364 B-MODE fetal ultrasound images were obtained from 247 pregnant women from April 07, 2014 to March 21, 2018 at the Evaluating Maternal and Fetal Markers of Adverse Placental Outcomes Clinic (EMMA) clinic at the British Columbia Women’s Hospital and Health Centre (BCW). The patients include 244 singleton and 3 twin pregnancies and gestational age range from 8 weeks to 34 weeks. The patients ranged from 17 to 50 years old and 40 patients were primigravida. The images were acquired with GE Voluson E8 (Boston, USA), GE Voluson E8 Expert (Boston, USA), Toshiba Aplio 500 (Minato, Japan) and Philips IU22 (Amsterdam, Netherlands) machines by 18 different sonographers.

The CNN was developed using the Keras framework. The CNN was trained using a NVIDIA Titan V GPU (Nvidia Corporation, Santa Clara, USA).

4.3.2 Design of Segmentation Method

A CNN, based on the existing network U-Net [82], was chosen to be applied to placenta segmentation as CNNs have shown promise recently in semantic segmentation problems [58]. The major advantage of a CNN compared to classical feature maps is that its deep architecture allows for the processing of a diverse range of feature maps to associate a combination of features to a specific classification. A CNN can characterize anatomy with features that are more complex than edge detec-
tion or statistical analysis of granularity. Although more computationally complex, the increased power of modern graphics processing units allow for CNNs to be usable for automated recognition of medical images. There are two major challenges identified specific to ultrasound images.

4.3.3 Limitations of Existing Networks

First, the ultrasound image may be manipulated by different operators and equipment. Operators may adjust time-gain compensation, alter speckle patterns changing frequency, or change the focus on the image. These modifications add additional complexity in image recognition, as a placenta may appear to have different granularity or brightness. Thus, placenta segmentation methods using a specific transducer and image settings may not be generalizable to other machines, decreasing the usability of such methods. To resolve this, the diverse dataset acquired was used to train a CNN to recognize placenta from different imaging scenarios.

Second, ultrasound contains unique imaging artifacts due to the interactions of acoustic waves. Acoustic shadows are particularly troublesome, as they occur when an acoustic wave reflects completely off a boundary between two significantly different materials [51]. The region below the boundary then appears completely dark. This commonly occurs in fetal ultrasound when the bones in the skull or spine obscure anatomy below. When shadow obscure portions of the placenta, its shape may seem irregular or divided into multiple sections. If an algorithm is trained to recognize continuous contours or shapes, it may miss a portion of placenta entirely due to being split by a shadow. If a network is trained to reject darker regions as non placenta, it may miss regions of placenta containing a cyst or an infarction, both of which appear dark. To address this, we modified the CNN architecture with an addition layer from an automated acoustic shadow detection algorithm to improve accuracy in detection with shadows present.

4.3.4 Convolutional Neural Network Design

The segmentation system consists of initial preprocessing and then the preprocessed image is used as an input to both an acoustic shadow detection step and a convolutional step in parallel. The two steps produce an output feature map for
Figure 4.1: Pipeline of the placenta detection method. After preprocessing an input image, the image is fed to two parallel paths. The first is through convolutional layers similar to U-Net to produce a preliminary placenta segmentation. The second is an automated shadow detection method to tune the optimization weights of the preliminary segmentation. After convergence of the optimization, a final output placenta map is produced.

For preprocessing to the network, first, all images are normalized and zero centered to reduce the sensitivity the loss function to parameters changes as activation functions used in the network are optimized for a zero-mean dataset.

The convolutional layers that preprocessed data are passed into resemble the U-Net architecture proposed by Ronneberger et al [82]. U-Net was chosen as U-Net provides sufficient simplicity compared to architectures requiring large intermediate layers, such as SegNet [6], or architectures requiring multiple residual pathways, such as ResNet [37]. The images are passed through a series of initial convolutional activation layers to recognize general features. The image is then...
Figure 4.2: Visualization of the convolutional layers. The input image is convolved with a series of weighted activation feature maps and the resulting activation output is compared to the ground truth to optimize intermediate weights such that the output of an activation pathway corresponds to either placenta or non-placenta. In this particular architecture, the feature maps are down-scaled with max-pooling to reduce computational complexity with the more detailed maps before being up-scaled again. Several hyperparameters and dropout pathways were adjusted to improve placenta segmentation.

down-scaled with max-pooling to improve computational efficiency in more specialized feature maps and then up-scaled produce a final segmentation map.

4.3.5 Parameter Selection

The training parameters were a learning rate of 0.0001, an epoch number of 100, a batch size of 1, a normal distribution for weight initialization, and a dropout regularization ratio of 0.50 on the two deepest layers. These parameters were chosen similar to those suggested by Ronneberger et al. [82] in the original proposal of U-Net and from similar studies implementing U-Net for ultrasound images [96] [42]. The dropout layers to randomly exclude data in each epoch to prevent overfitting.

The gradient descent optimizer was not selected as ultrasound images contain substantial noise compared to modalities with lower granularity, such as CT or MRI. This results in various local minima in the $n$-dimensional space of edge weights. To
address this, a momentum term is required such that traversing the \( n \)-dimensional space is not at a constant step size, but a step size that varies and increases to sample sets of weights outside of a discovered minimum. This reduces the likelihood of convergence in a local minima. The optimizer in particular is known as an adaptive momentum estimation (ADAM) optimizer \([50]\) as described in Equation 4.1

\[
m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t
\]

\[
v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2
\]

Where \( m \) and \( v \) are moving averages at iteration \( t \), \( g \) is the gradient of the current epoch, and \( \beta \) is a momentum parameter. \( m \) and \( v \) are estimates of the mean and variance of the gradients. They are first initialized to 0, which results in the iterations to be biased towards zero during the first iterations. To address this, a bias correction term is applied, as seen in Equation 4.3

\[
\hat{m}_t = \frac{m_t}{1 - \beta_1^t}
\]

\[
\hat{v}_t = \frac{v_t}{1 - \beta_2^t}
\]

The biased corrected mean and variance estimates \( \hat{m}_t \) and \( \hat{v}_t \) are then used to update the set of weights \( w \) in Equation 4.5

\[
w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}
\]

Where \( \eta \) is a step size parameter and \( \varepsilon \) is a smoothing term parameter.

To detect acoustic shadows, a the B-MODE shadow detection method from Chapter 3 reference was used. There was no access to RF data for speckle analysis. The map of detected acoustic shadows is then used to identify pixels of shadowing on the initial segmentation map output of the convolutional layers to double the convolutional weights of the shadow regions. The modification of the weights follows the intuition of attention networks in literature, where a specific set
of features is considered unique to the cost function [102]. In this study, the specific feature is the region of acoustic shadow and the uniqueness is that it may split a placenta to two segments and create a region of unexpected textural information. This causes a problem if the network is trained to recognize specific geometric features of non-shadows images for placenta segmentation, such as attempting to identify an object with a single closed contour.

With the weights doubled, a placenta with shadow is delineated from patterns associated strictly with non-shadow images. As a result, the back-propagation and optimization of previous convolutional weights will be processed differently for images containing acoustic shadow. Geometrically, if $N$ weights are visualized as parameters in $N$-dimensional space with some local minima, then the weight modification forces the optimization to be driven away from the current local minima and search for a local minima that is optimized for images with acoustic shadows. After optimization and as the network converges, an output placenta map is then computed.

Selection of the extent to modify the convolutional weights was performed empirically, where a weight change multiplier on the order of 1.4-3.5 was observed to increase the mean Dice coefficient accuracy above 0.85 for shadow images.Doubling the weights by a static factor of 2 is likely not the optimal factor for each batch of data. Future studies are required to develop dynamic methods to adjust weights during training the are specific to each data set. This may include, for instance, a separate parallel network that optimizes the weight modification factor to increase the prediction accuracy, similar to adaptive parameter prediction networks used in literature for photograph object classification [68].

Another method to address the inaccuracy of shadow images is to add more training data. The optimization process may potentially ignore features specific only to non-shadow images. However, this likely requires more data that the dataset acquired for this clinical study as there needs to be sufficient data with acoustic shadows, particularly when acoustic shadows occur from different angles and depths.
4.3.6 Validation

To create ground truth images for the neural network, annotators were trained by an obstetrician to manually segment the placenta from all images.

To train and test the CNN, the set of 1364 images were randomized and separated, with 70% used for training, 15% used for validation during training, and 15% used for final testing after the model has been created (205 images). The segmentations of the final testing was compared to manual segmentation for accuracy. A second test was conducted to test the same model only on images from the full dataset that contain acoustic shadows (76 images). This test was to emphasize accuracy improvements with the additional acoustic shadow weight layer. A Ljung-Box Q-test was performed to test the autocorrelation of the Dice coefficient values.

To compare accuracy of the neural network, other common architectures were implemented, namely SegNet [6], E-Net [70], and U-net without a shadow detection layer. To evaluate inter-observer accuracy, a subset of 100 images were segmented 5 times by the annotators and accuracies were compared against segmentations of the same image.

4.4 Results

Accuracies of all the architectures are summarized in Table 4.1 for all images and for the subset of images containing acoustic shadows. Selected images comparing manual segmentation and automated segmentation are shown in Figure 4.3. Worst case images are shown in Figure 4.4. A comparison of manual segmentations achieved a Dice coefficient of 0.93 ± 0.03 and 0.91 ± 0.04 on all images and on the acoustic shadow subset.

The proposed method performs with higher accuracy than previous studies (p < 0.05 for all images and p < 0.05 for images with acoustic shadows) and similar to the Dice coefficient obtained in manual segmentation within the uncertainty of inter-observer variability of ±0.03 for all images and ±0.04 for images with acoustic shadows. The proposed method performs more accurately for images containing acoustic shadows than architectures without ultrasound-specific shadow detection. This is expected as shadows result in a region resembling a large loss.
Table 4.1: Dice similarity coefficients when comparing different detection methods to manual segmentation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Dice Coefficient</th>
<th>Acoustic Shadow Images (76 Images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Images (205 Images)</td>
<td></td>
</tr>
<tr>
<td>SegNet</td>
<td>0.90±0.06</td>
<td>0.74±0.04</td>
</tr>
<tr>
<td>E-Net</td>
<td>0.86±0.04</td>
<td>0.67±0.03</td>
</tr>
<tr>
<td>ResNet</td>
<td>0.89±0.04</td>
<td>0.71±0.05</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.91±0.03</td>
<td>0.75±0.05</td>
</tr>
<tr>
<td>U-Net + Shadow Layer</td>
<td><strong>0.92±0.04</strong></td>
<td><strong>0.87±0.04</strong></td>
</tr>
</tbody>
</table>

Figure 4.3: Output segmentation maps compared with manual segmentation. Three cases are presented including one of a placenta in plain view (a,d,g), one of a placenta in partial view (b,e,h), and one of a placenta with an acoustic shadow (c,f,i). The focus of the method is to handle cases with acoustic shadows, where the detection is able to identify placenta regions even though they are not a single closed contour.
in signal and activation maps that attempt to associate features in this region may be unable to classify the stark change in echotexture between a placenta and an obscuring shadow.

The proposed method achieves a higher Dice coefficient (0.92 ± 0.04) than previous placenta segmentation studies. One study reported a Dice coefficient of 0.87 ± 0.13 [91] for a random walker method and another achieved a Dice coefficient of 0.73 ± 0.07 utilizing the open-source CNN DeepMedic [59]. In addition, the proposed method was developed with a dataset of 1364 images, greater than the largest dataset in past studies, which had 300 images. The increased computational efficiency of the proposed architecture, aided by shadow detection and a
varied training data set likely resulted in higher segmentation accuracy.

4.5 Discussion

The segmentation method achieved higher accuracy than previous studies, is automated, and requires no user input beyond the acoustic shadow pulse length parameter. The results support the potential for the segmentation method to be used to process placenta ultrasound methods to aid further analysis of only the placenta without interference from other portions of a fetal ultrasound.

An important result of the experiment is the validation of the proposed method over a varied dataset. The data collected was obtained by multiple sonographers with different ultrasound machines, and image settings. This provides support for the usability of the method. A versatile solution may address the motivation of automated placenta detection for novice or point-of-care users with different clinics and users.

Ultimately, the goal of developing a placenta segmentation tool is to provide an input to a placenta health assessment tool. With successful placenta segmentation, the next step would be acquire a set of placenta images with clinical outcomes, such as whether the placenta has pre-eclampsia or if a physician assessed the placenta as normal. After placenta segmentation is performed on the images, texture analysis on the placenta, such as with a CNN, can be performed to correlate certain textures to a classification. The placenta assessment tool can then be packaged into software where clinicians can upload placental ultrasound images and receive a risk probability for placental abnormalities.

4.6 Limitations

Several inaccuracies occurred when the neural network misidentified non-placenta regions as placenta. The most common source of false positives was segmenting the uterine wall as part of the placenta. The uterine wall is normally in contact with the placenta, indicating that the detection has difficulty in differentiating placenta and non-placenta regions near the placenta boundary. This may be due to the placenta boundary having slightly different echotexture in a thin region. As the placenta boundary areas are small compared to the placenta body, it is expected...
that ML is more successful in recognizing the echotexture of the placenta body due to the ample amount of image data and less successful in identifying boundaries. This may be addressed in future studies by characterizing different anatomy with a similar training approach to provide more classifications beyond placenta and non-placenta. If the dataset contains training labels for the uterine wall, fetal head, or amniotic fluid, for instance, a ML algorithm may characterize boundary characteristics of other anatomy to recognize non-placental areas in more detail.

Another limitation of this work was that the placenta images were taken by trained fetal sonographers. Thus, all images that were recorded were considered to be of acceptable quality for visual analysis by a physician. The dataset is not reflective of images taken by users who have not received training to take fetal ultrasound images, such as point-of-care users. The method is then not fully validated on images that may not show the placenta as clearly, as the detection algorithm has not been trained on such images. In the future, to increase the generalizability of the CNN, more training data that includes images taken by non-experts can improve the training phase of the CNN.

Further work in placenta segmentation would include integrating placenta detection into existing clinical scanning procedures to develop real-time detection. This may include integration into an ultrasound machine to extract image data and present a segmentation of the placenta at speeds similar to the frame rate. This requires further testing and optimization to be usable while a scan is being performed and the ultrasound image display is constantly updated.

4.7 Conclusion

In this chapter, we proposed an automated placenta segmentation method by modifying an existing CNN with acoustic shadow weights, which contribute to reducing errors due to unique ultrasound artifacts. The work combines recent ML methods with classic image analysis based on ultrasound physics of acoustic shadows. The method has achieved detection accuracy comparable to manual segmentation over a large dataset of 1364 images obtained over several years with different operators and equipment. The results provide support that automated placenta segmentation can be performed for a variety of images and applicable for identification of
the placenta for non-specialized users or to automatically extract the placenta for large-scale data processing applications, such as ML to classify abnormal placentas.
Chapter 5

Binary Classification of Placenta Ultrasound Images

5.1 Introduction
The main objective of placenta image analysis is to aid a physician in assessing the placenta ultrasound which can be done non-invasively throughout the pregnancy. There are several perceived limitations in achieving this goal from reviewing the literature. This section first discusses the challenges of placenta classification and constraints due to the clinical study. Next, this section discusses the scope of the work, the CNN developed, the results from this preliminary study, and a discussion of limitations and future directions.

5.1.1 Challenges of Placenta Ultrasound Classification
An automated placenta ultrasound classification tool can potentially provide the benefits of automated analysis to assist physicians, potential frequent screening with accessible ultrasound, and early detection if the tool can identify abnormalities in a first-trimester ultrasound as described in Chapter 2. This is a typical ML classification task, where a ML model predicts a label (disease or no disease) for the dataset of a single patient (patient ultrasound images and charts). The model would attempt to correlate complex features in the dataset to the label. In placenta
imaging, there is the difficulty in that the correlation between complex features and a clinical outcome of disease or no disease is not fully understood. It is not fully known what appearance changes in ultrasound are related to preeclampsia or IUGR \[8\]. An ultrasound image can also vary in appearance from different angles, machine settings, and operators, which may affect the appearance of the placenta. Due to the invasive nature of pathological studies with biopsies, there is limited comparisons of a placenta ultrasound taken with a pathological micrograph at the same gestational age to identify what a certain pathological feature looks like on ultrasound \[88\]. Thus, there is limited understanding as to what kind of features that a ML model should look for or is the current data available contains sufficient data and variability to train a ML model.

The ethics approval in the clinical study pursued in Chapter 4 allowed additional retrieval of retrospective data of placenta provided ultrasound placental sweeps for 367 patients with retrospective data of patient risk factors, assessment report from the EMMA clinic, and reported diagnosis of preeclampsia or IUGR. A full end-to-end placenta classification tool would receive numerical patient risk factor data and ultrasound sweep as inputs to produce a disease or no-disease classification as an output to correlate with an unhealthy outcome such as preeclampsia or IUGR. However, machine learning techniques require a large amount of training data. With patient risk factor data from 367 patients only 64 diagnoses with preeclampsia or IUGR, the risk factor dataset may be insufficient to construct a generalizable model.

5.1.2 Scope of Work

The scope of the work in this chapter was motivated by the challenges in placenta classification in two ways. First, due to limited patient chart data, the methods focuses on utilizing only the placenta appearance in ultrasound data. Second, due to the unknown feasibility of a model classifying clinical outcome based on the ultrasound, the main task investigated was an intermediate step to classify whether a placenta had a normal or abnormal appearance, which is similar to the existing workflow at the EMMA clinic.

There is uncertainty as to how a placenta ultrasound appearance is correlated
with preeclampsia or IUGR, which poses a challenge to develop a model to classify a placenta appearance as associated to a certain disease [104]. At the EMMA clinic, the current standard is that a physician visually inspects the ultrasound images and notes whether the placenta appears normal or abnormal depending the placenta homogeneity, presence of lesions, and size of placenta. Depending on the assessment, the patient is then stratified into a risk category. In the physician assessment, though it is subjective, there is a known correlation between the ultrasound image and the placenta abnormality classification.

Due to the unknown feasibility to predict clinical outcome from placenta appearance, this thesis attempts two tasks. The first task is to classify a placenta as normal or abnormal appearance. The second, which is exploratory at this stage, is to classify a placenta as associated with disease or no disease. In both cases, the tool would still achieve the benefits of being automated, usable early, usable frequently, and accessible by non-experts.

We developed two models from the same CNN architecture. Both models utilize ultrasound data to classify a placenta. The first model to predict whether a physician manually assessed the placenta as normal or abnormal was named as the ultrasound assessment classifier. The second model to predict whether the placenta will result in either preeclampsia or IUGR was named as the clinical outcome classifier.

5.2 Related Work

There is limited existing work in automated placenta early classification of visible abnormalities, likely due to two reasons. The first is that study of the placenta is relatively underinvestigated as placenta ultrasound assessment guidelines discussed in Chapter 2 are limited to qualitative metrics, which are subjective and may be inconsistent across observers. The second is that only recently have complex machine learning models, such as neural networks, gained popularity due to the increased processing power of GPUs. There several more studies on pathological examinations of an ex-vivo placenta to correlate lesions with adverse fetal outcome [72], though this would not be applicable for early disease prediction before the placenta is can be physically examined by a pathologist.
Although there are many studies with automated ultrasound analysis in other domains, three studies (\cite{56}, \cite{20}, \cite{18}) were selected to be especially relevant to the motivation of this thesis. The studies focus on placenta analysis, clinical outcome prediction in ultrasound for a specific disease with evident visual appearance changes, and clinical outcome prediction in ultrasound for a set of diseases with no well-known evident visual appearance changes. The third study had the most similar tasks to the objectives in this thesis, despite with different anatomy. Each study provides proof of concept to techniques that may be applicable to placenta classification.

In the first study, Li et al. \cite{56} developed a model for automated grading of placenta maturity with a discriminative model using clustering techniques for higher order statistics of a placenta ultrasound image. A sensitivity of 98.04\% and a specificity of 93.75\% were reported with this method with 443 placental images. However, the study was limited to one ultrasound transducer, with scans acquired in a standard fashion. The discriminative learning model provides support that automated learning of a placenta ultrasound image can correlate texture features with classification, which is a goal of diagnostic placenta classification. The classification of placenta maturity is with standardized metrics such as placenta size and a more complex set of feature may be needed to classify a placenta abnormalities, which are not as well-defined as maturity.

Although placenta ultrasound classification methods are limited, there are many diagnostic ultrasound classification applications with other diseases utilizing neural networks that attempt to achieve similar goals to this thesis. In the second study, Chi et al. \cite{20} utilized the common CNN architecture GoogLeNet to classify thyroid ultrasound nodules as either malignant or benign, with a classification accuracy of 98.29\% on 428 images. This aims to achieve a similar goal as placenta classification and the results indicate that neural networks have the potential to analyze ultrasound images. In the case of placenta classification, placenta texture is not a single anatomical feature like a thyroid nodule, but is rather relevant to the entire placenta which may exhibit many different pathologies. Thus, placenta classification poses a different problem in that either preeclampsia or IUGR are considered abnormal potential outcomes.

In the third study, Cheng et al. \cite{18} utilized another common CNN, VGGNet,
to classify whether an abdominal ultrasound is normal or abnormal, with abnormalities defined as predicting a patient exhibiting one of six abdominal conditions. Investigating abdominal abnormalities is not as well defined as investigated localized thyroid nodules, as there was no reported definition of how ultrasound texture corresponds to a specific disease. This classification task is relevant to this thesis as it is more similar to placenta classification where abnormality is a generalized outcome. The study reported an accuracy of 77.9% using a total of 1423 images. The decreased accuracy compared to thyroid nodule classification in [20] is expected, as malignant thyroid nodules have more evident visual changes, such as size and irregular thyroid shape, compared to an abnormal abdominal ultrasound. This study motivates further exploration to outcome prediction, though the architecture selected may not have enough complex feature relationships.

The above related studies indicate that there is potential to investigate ML in placenta ultrasound images to correlate detailed textures with a diagnostic classification. At the same time, the studies indicate that there are significant challenges in classification of abnormalities from general ultrasound scans compared to specific classification based on known physical features.

5.3 Materials and Methods

5.3.1 Data Collection

The collection of ultrasound images was approved by the University of British Columbia Clinical Research Ethics Board (ID: H18-01199). A total of 7831 frames of ultrasound images taken from 367 placenta ultrasound were obtained from retrospective data at the EMMA clinic. The sweeps correspond to 367 patients from April 07, 2014 to March 21, 2018. In addition, charts of patient risk factor data, the reported ultrasound placenta assessment recorded by the physician, and diagnostic outcome of the patients were obtained. The images were acquired with GE Voluson E8 (Boston, USA), GE Voluson E8 Expert (Boston, USA), Toshiba Aplio 500 (Minato, Japan) and Philips IU22 (Amsterdam, Netherlands) machines by 18 different sonographers.

The CNN was developed using the Keras framework. The CNN was trained
using a NVIDIA Titan V GPU (Nvidia Corporation, Santa Clara, USA).

5.3.2 Convolutional Neural Network Design

Image classification has been a popular problem for methods using CNNs and several existing architectures have been developed for classification motivated by competitions such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [84], such as ResNet [37], AlexNet [52], and VGGNet [90]. The study by [18] used the CNN VGGNet to classify abnormal abdominal ultrasounds, indicating that the popular CNNs that performed well at ILSVRC may potentially be useful for ultrasound classification.

ResNet was selected from its success at ILSVRC, achieving higher accuracy that VGGNet at ILSVRC 2015 and providing deep architecture with higher complexity than the neural network used in placenta segmentation in Chapter 4. Of all architectures at each annual ILSVRC, ResNet contained some of the highest feature complexity [57]. In the task of placenta classification, where the current standard of visual analysis does not have well defined metrics for evaluating images and ultrasound images may contain detailed information in speckle, greater feature complexity was desired to explore complex details beyond the ability of visual analysis. The Resnet-50 variant was selected as it provides greater speed than the ResNet-101 or ResNet-152 variants with similar accuracies from the ILSVRC standard tests [100].

The key design component of ResNet is the application of residual layers. This is motivated by the observation that continuously stacking convolutional layers results in a decreased training error and test error, reducing the capability of exploring higher complexity features. An overly complex network may be expected to overfit data, resulting in a lower training error and higher test error. However, it was observed by He et al. [37] that the training error also increases, which leads to the hypothesis that the error is not solely overfitting, but that deeper models are more difficult to optimize.

The solution applied was that the output $H(x)$ of a deeper layer $L_i$ is converted to a residual $F(x)$. The residual $F(x)$ is the value of $(H(x) - x)$, where $x$ is the input of a previous convolutional layer. Intuitively, instead of learning to produce
Figure 5.1: A visualization of the ResNet architecture. Only two residual blocks are displayed for clarity. The full architecture used contained 50 layers. ResNet is a much deeper network than U-Net used in placenta segmentation, allowing more complex details to be activated. The residual blocks provided references of previous layers to initialize the output of a deeper layer to assist optimization.

an output $H(x)$ which requires high complexity optimization, the network instead learns what is required to be incrementally added to the input $x$ to achieve a similar output. In this way, the output is initialized to the identity of $x$ and optimization complexity is decreased if the optimized output value is close to $x$. Note that this hypothesis was tested experimentally in ILSVRC and there is limited rigorous mathematical proof for this concept at the moment. This output-input relationship between convolutional layers is called a residual and is visualized in Figure 5.1.

ResNet-50 was combined with the existing pipeline for acoustic shadow detection and placenta segmentation described in Chapter 3 and Chapter 4. The full classification system was tested on two classification tasks: classifying a placenta as normal or abnormal and classifying a placenta as being associated with no placental disease or placental disease. The data was separated with 80% belonging to a training set and 20% belonging to a final testing set, similar to [18]. All reported accuracies were when testing on the testing set. The proportion of testing data was increased compared to segmentation as it was not known what exact correlation there is with placenta images and abnormalities. Hence, a larger testing set was used to prevent overfitting the training data.

128 epochs were used in training, with a learning rate of 0.001, a batch size of 1, and a dropout regularization ratio of 0.50 was used every in the deepest layer to
The ADAM optimizer was chosen for similar reason as discussed in Chapter 3. The hyper-parameters were similar to previous studies [18], with the maximum number of epochs and batch size that does not result in insufficient memory allocation with resources used.

## 5.3.3 Validation

To validate the task of classifying a placenta as abnormal or normal appearance, the ground truth was obtained from retrospective clinical chart. The clinical chart data contains a ”placenta assessment” flag. When the patient has visited the EMMA clinic for an ultrasound exam, a physician assesses the placenta ultrasound. This corresponded to 1719 abnormal images and 6112 normal images. The placenta is flagged as abnormal if any one of the following are observed:

1. If two of the following are observed: placental thickness greater than 5.0 cm, placental length less than 11.0 cm, jelly-like echotexture, or greater than three infarcts.

2. The uterine artery Doppler profile shows a bilateral notch as previous studies indicated bilateral notching as a potential risk factor for preeclampsia [30].

3. The uterine artery Doppler profile shows a pulsatility index greater than 1.55, from previous studies indicating potential abnormal blood flow [64].

To validate the second exploratory task of classifying a placenta as associated with either placental disease or no placental disease, the ground truth was obtained from the clinical chart recorded after delivery. After delivery, a physician notes if the patient was diagnosed with preeclampsia or IUGR. For each ultrasound image, the ground truth was defined as with disease if the patient was diagnosed with preeclampsia or IUGR and without disease if the patient was not diagnosed with either. This corresponded to 1638 disease and 6193 non-disease images.

## 5.4 Results

The confusion matrix of the placenta ultrasound assessment classifier is shown in Figure 5.2 and the evaluation metrics are summarized in Table 5.1. The placenta
Table 5.1: The recorded sensitivity and specificity of the placenta ultrasound assessment classifier.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placenta Ultrasound Assessment Classifier</td>
<td>0.87</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 5.2: The recorded sensitivity and specificity of the clinical outcome classifier.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Outcome Classifier</td>
<td>0.60</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The placenta ultrasound assessment classifier achieved a sensitivity of 0.87 and a specificity of 0.91. The confusion matrix of the clinical outcome classifier is shown in Figure 5.3 and the evaluation metrics are summarized in Table 5.2.

The placenta ultrasound assessment classifier achieved a sensitivity of 0.87 and a specificity of 0.91. The clinical outcome classifier performed worse, with a sensitivity of 0.60 and a specificity of 0.81.

To visualize the convolutional layers, an example of the outputs of the first 6 convolutional kernels of the ultrasound assessment classifier are displayed in Figure 5.4. The output represents the resulting matrix when the input image is convolved with the optimized convolutional kernels in the CNN, representing a region that has been extracted to highlight important features for the layer to process.

To visualize the segments of the placenta that the CNN focuses on to classify the image, the class activation map (CAM)s of a normal and abnormal placenta are displayed in Figure 5.5. The CAMs represent regions of the image that are activated to have nonzero convolutional output to pass onto deeper layers. The significance of these regions is that they are the major features that a CNN has identified to have an impact on its classification decision of an image.
Confusion Matrix for Placenta Ultrasound Assessment Classifier

Ground Truth Classification

<table>
<thead>
<tr>
<th></th>
<th>Abnormal Appearance</th>
<th>Normal Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal Appearance</td>
<td>279</td>
<td>106</td>
</tr>
<tr>
<td>Normal Appearance</td>
<td>42</td>
<td>1123</td>
</tr>
</tbody>
</table>

Figure 5.2: The confusion matrix of the classifier to predict if a placenta had been assessed as normal or abnormal appearance from the ultrasound image using the ultrasound assessment classifier.

5.5 Discussion

The result indicate that the placenta ultrasound assessment classifier performs with reasonable accuracy, especially compared to the clinical outcome classifier. This is expected as the placenta ultrasound assessment attempts to correlate the appearance of a placenta as either normal or abnormal with a ground truth mainly based on physician assessment of visual features. The clinical outcome classifier attempts to correlate features in an ultrasound image to predict a diagnosis of either preeclampsia or IUGR. This is a significantly more difficult problem, as the two conditions are not defined by ultrasound appearance but rather physiological conditions.

The hypothesis for the clinical outcome classifier was that the structural changes discussed in Chapter 2 would present in ultrasound images. The results indicate that the hypothesis is inconclusive as the network was unable to achieve high prediction accuracy utilizing the ultrasound images. However, this does not dispel the
### Confusion Matrix for Clinical Outcome Classifier

#### Ground Truth Classification

<table>
<thead>
<tr>
<th></th>
<th>Placental Disease</th>
<th>No Placental Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Placental Disease</strong></td>
<td>240</td>
<td>216</td>
</tr>
<tr>
<td><strong>No Placental Disease</strong></td>
<td>146</td>
<td>948</td>
</tr>
</tbody>
</table>

**Figure 5.3:** The confusion matrix of the classifier to predict if a placenta had been assessed as being associated with disease or no disease using the clinical outcome classifier.

The potential of utilization a CNN to achieve clinical outcome classification as there are several limitations to this study and method which may be addressable in the future.

The placenta ultrasound assessment classifier performed with accuracy in a similar range to similar studies [18], indicating that the CNN can potentially achieve the goals of placental ultrasound assessment. The potential generality of this method is supported by the dataset, where ultrasound images came from different patients with different machines and operators. The implication of this is that the placenta ultrasound assessment classifier may be packaged as a software tool in the future to assist physicians in recognizing an abnormal placenta, to provide automated analysis, early analysis, and access to non-specialists by utilizing portable ultrasound.

The development of this placenta abnormality assessment tool has the poten-
Figure 5.4: The class activation maps of a) normal placenta and a b) abnormal placenta from the ultrasound assessment classifier. The activation maps show the segments of the placenta that provided the most important features used to discriminate a placenta as normal or abnormal. Visually, there is no recognizable pattern with the location of the activations. This indicates that the activations are not location-specific and that features throughout the entire placenta was used to correlate with an abnormal or normal prediction. This may include details such as small lesions or inhomogeneities in echotexture that are not visible to the human eye, unless magnified. This map may aid physicians in identifying regions of interest for pathological study of the placenta.

The CNN classifies the placenta based on the features of an ultrasound image. Once a classification is made, we can possibly use the CNN to construct a CAM to visualize the location of the neuronal activations. The pathologist can then identify regions of interest for histological study, as the location of the activations for a placenta classified as abnormal indicate that the CNN has observed differing features compared to a normal placenta. In addition, we can observe the output of the convolutional kernels to visualize what features are being extracted, although this requires further computational study as the features have nuanced details that are difficult to consistently analyze visually. Examples of a CAM and output of convolutional kernels are visualized in Figure 5.4 and Figure 5.5.
Figure 5.5: The output of the first 6 convolutional kernels. From the output, a physician can visually observe what sections and what type of features are being extracted from the image, although it is difficult to quantify without detailed numerical analysis.

5.6 Limitations

There are several limitations associated with the data and the method of utilizing a CNN on the placental sweeps. The low accuracy for the clinical outcome classifier indicates that there are major limitations in the current model and is likely not usable in a clinical setting at the moment.

First, the ground truth for placental assessment classification is the retrospective physician assessment. This is a subjective assessment based on metric discussed in Chapter 2. Thus, the performance of the placental assessment classification tool is limited to matching manual analysis. Although the tool still provides benefits of automation, accessibility, and ease of use, it does not provide additional diagnostic accuracy to improve the standard of practice.
Second, the training for both classifiers was performed on placental sweeps. This contains continuous video-based frames of the placenta, with adjacent frames containing similar images. The ground truth, where a patient received either a normal or abnormal assessment, is a flag that is associated with the entire sweep. If an ultrasound image contains a portion of the placenta that appears normal, it would still be flagged as abnormal if the physician has noted an overall abnormality of the sweep. This provides false labels for several frames. The placenta ultrasound assessment classifier was still able to identify features, likely because there is a sufficient interval of frames containing abnormal features to train the network. To improve this, each individual frame may require expert classification as to whether the placenta in the image is normal or abnormal appearance. An alternate method is to consider all the frames in the sweep as a single data point and stitch the frames together as a 3D-like dataset. This decreases the size of the set available for training. With the relatively small dataset we obtained, it was determined that this was unfeasible for machine learning methods that require a large training dataset.

Third, there is likely insufficient data for training a clinical outcome classifier. As the limitation of adjacent frames containing similar image may have affected the placental ultrasound assessment classifier, it would likely affect a clinical outcome classifier to a greater extent. Classifying clinical outcome is more nuanced, as there is no clearly known correlation between the texture of a placenta to placental disease. Thus, accurate labels are more important. Additional data, such as recorded patient risk factors that are suspected to be correlated with placental disease, would likely increase the predictive power of a classifier that solely uses ultrasound images. However, with the current dataset, there is an insufficient number of patients to train a generalizable classifier. With more data in the future, clinical patient data can be concatenated with ultrasound image data as a single input vector to a CNN. The CNN and resulting CAM can then determine which risk factors and portions of the images are the most relevant to placental disease.

5.7 Conclusion

In this chapter, we proposed a CNN architecture to classify a segmented placenta image. The classifier was tested on two different tasks. The first task was to clas-
sify the placenta as abnormal or normal appearance, with recorded visual analysis by a physician as the gold standard. The second task to classify the placenta as associated with disease or no disease, with the recorded clinical diagnosis of the patient as the gold standard. The tasks were tested on retrospective ultrasound data of 7831 images from 367 patients. The former placental ultrasound assessment task achieved a sensitivity of 0.87 and a specificity of 0.91. The latter task performed with a sensitivity of 0.60 and a specificity of 0.81. There are several limitations with the dataset, mainly that the images were still frames of placental sweeps while the ground truth labels were for the entire sweep. Likely, more data is required to evaluate and develop a full end-to-end placenta clinical outcome classification tool. Although classification of clinical outcome was not successful, classification of whether a placenta was normal or abnormal, based on clinical guidelines, performed better. The accuracy performance of the placenta ultrasound assessment tool indicates that the CNN developed can still provide value for automated, accessible, and early screening of an abnormal placenta. The results show the potential of ML analysis of the placenta, as complex texture features were used to complete a classification tasks. The future development of ML methods may be improved to perform more complex tasks, such as predicting a diagnostic outcome.
Chapter 6

Conclusion and Future Work

6.1 Summary of Contributions

The work conducted in this thesis consists of conducting two clinical studies and developing three major components for a placenta ultrasound classification tool. This section summarizes the contributions by the author.

6.1.1 Contributions in Developing a Placenta Analysis Pipeline

The first contribution component was the development of an automated acoustic shadow detection method. The method was able to detect acoustic shadows on both RF and B-MODE images, achieving a Dice coefficient of $0.90 \pm 0.07$ and $0.87 \pm$ respectively when compared to manual segmentation. The author has developed the software and processing tools for both methods. The author has developed a RF speckle statistics analysis algorithm and improved on previous B-MODE pixel entropy analysis by reducing all detection methods to a single parameter. The author designed and conducted a clinical study to scan 37 human volunteers, obtaining 222 ultrasound images of the forearm, elbow, and ribcage. The method developed achieves accuracy similar to manual segmentation, is usable by non-experts with its single-parameter operation, and is capable of detecting shadows from different scenarios. In addition to detecting shadows, the author has also contributed an analysis of the RF speckle to investigate the statistical properties of speckle in shadows and non-shadows, providing insight to the ultrasound scattering physics
of acoustic shadowing.

The second contribution was the development of an automated placenta segmentation method utilizing a modified CNN. The method was developed as a second stage in the placenta classification pipeline with the acoustic shadow detection from the first contribution as an input. The author designed and conducted a clinical study to retrieve 1364 retrospective fetal ultrasound images from 257 patients to test the method, achieving a Dice coefficient of $0.92 \pm 0.04$ for all images and $0.87 \pm 0.04$ for images containing acoustic shadow. The segmentation method outperformed past segmentation studies and existing segmentation architectures.

The author has developed all the software to segment the placenta, with the CNN being derived from the existing architecture U-net [82]. The method developed achieves accuracy similar to manual segmentation, is automatic, and was tested on ultrasound data from different patients, machines, and operators.

The third contribution was the development of a placenta classification method, utilizing a CNN with residual connections. The method was developed as the final stage in the placenta classification pipeline, with placenta segmentation utilizing the first two contributions as an input. The author continued the clinical study from the second contribution to obtain 7831 fetal ultrasound images from 367 patients to validate the method. The method was capable of classifying the placenta as abnormal or normal, as defined by the physician assessment of the ultrasound images in patient records, with a sensitivity of 0.91 and a specificity of 0.87. The method was unable, however, to predict placenta to be diagnosed with placental disease with similar accuracy. The author developed the software of the residual CNN, with the architecture design derived from the existing network ResNet [37]. The CAM metrics were added to the pipeline, to provide a visualization of class activations when a placenta was classified as normal or abnormal. The CAM allows users to locate the specific region of interest where the CNN detects an identifying feature of abnormality.

6.1.2 Contributions from Evaluation and Implications of Results

The evaluation of the placenta classification pipeline resulted in the author identifying implications of the methods developed. The following implications were
contributed:

1. **Demonstrating the potential of automated acoustic shadow detection:**
   The acoustic shadow detection method achieved accuracies similar to manual detection, indicating that the method has the potential to detect shadow in an automated way. Previous studies did not develop an easily usable system, often requiring multiple parameter tuning by experts.

2. **Demonstrating the capability of ML to identify a placenta in fetal ultrasound:** The performance of the placenta segmentation method supports the potential for using ML to automatically extract a placenta from an ultrasound image. The results indicate that complex texture details can be identified by ML to accomplish visual recognition tasks in fetal ultrasound. This can motivate future studies to improve on placenta segmentation by addressing the limitations identified, including expanding segmentation to identify subcomponents of a placenta and providing more data to increase the generalizability of the method.

3. **Demonstrating the potential of ML for future studies in placenta classification:** The methods developed, using ultrasound physics modeling and ML, has the potential to provide quantifiable analysis of placenta ultrasound images. The current standard of practice is visual analysis from a trained expert, which has limitations of subjectivity and lack of access. The results from this thesis indicate that methods such as ML can achieve tasks with classification accuracy in range of previous related studies in ultrasound analysis [18]. Although there are limitations to the method used, the accuracies can motivate future research in developing further placenta analysis methods as ML methods can process far greater image detail than human visual analysis.

4. **Understanding limitations of placental imaging for quantifiable analysis:** The sources of limitations of the methods, especially in classifying a diagnostic outcome, reveal several requirements to develop a high-accuracy placenta classification tool for routine clinical use. With ML, a large training dataset is needed for the model to be generalizable. Future studies on de-
veloping placenta analysis methods would require the retrieval of more data, including both ultrasound and numerical patient chart data.

5. **Understanding limitations of defining acoustic shadows in ultrasound:**
   The acoustic shadow detection analysis reveal that acoustic shadows were not a sharp change from high to low intensity on a B-MODE image, but rather a gradient in a transition region. There is currently a qualitative definition of an acoustic shadow to define where a shadow begins. Manual detection by an expert was used as a gold standard, though intra-operator errors often occurred near the boundary of the shadow. The results in the thesis indicate the potential for future studies to further investigate the statistics of an acoustic shadow, which may aid the development of a quantitative definition for consistent detection of a shadow region.

6. **Identifying the extent which the methods can serve as a physician aid:**
   The limitations of the placenta classification pipeline, including the low accuracy and potential to redesign the model with more data, indicates that the current method requires improvement to achieve the capability to classify a placenta to a clinical outcome. The results indicate that the tool can perform placenta abnormality assessment in an ultrasound similar to a physician. In its current iteration, the placenta classification method is likely to serve as an additional aid to a physician in analyzing a placenta, with benefits of automation or providing additional data for analysis. Due to the limited data, the method likely requires further training with additional data to achieve greater generalizability to the full population.

7. **Identifying the Ambiguity in the Class Activation Map:**
   The CAM provided a visualization of the regions detected by the CNN to be important when discerning a normal placenta compared to an abnormal placenta. Visually, the CAM showed no obvious pattern. This finding indicates that the regions correlated with abnormality are not easily analyzed with manual analysis. To explore the regions of interest, future studies are required. The ability for the CNN to achieve similar sensitivity to previous ultrasound abnormality detection methods indicates that class-activated regions contained useful
features for classification. The result support further development of numerical methods to understand a quantitative relationship between the activation maps and the physical structure of the placenta, possibly with histological studies in the future.

6.2 Limitations and Future Work

The limitations of each method are discussed in detailed in their respective chapters. This section summarizes the limitations and resulting future work possible to address them.

Limitations of acoustic shadow detection are that:

1. Acoustic shadows occur in a variety of anatomy. Development of a full universal acoustic shadow system requires testing on shadows produced by more combinations of equipment, operators, and anatomy in the study. Future studies could involve more extensive investigation, such as with a parameter grid search, to improve the generalizability of the method.

2. Reverberation artifacts can lead to a false shadow as reverberation from bone also presents as a bright region followed by a dark region below. To improve the acoustic shadow detection, integration of reverberation identification is required. This may be done in future work by utilizing the echo time information of the RF signals to determine if a pulse is due to an anatomical interaction of reverberation errors [99].

3. The speckle distributions of ultrasound images have been observed to vary at frequencies higher than 36MHz [23]. This would affect the binary threshold used to classify a region as shadow or non shadow if the speckle characteristics are not constant. Future work should investigate utilizing ultrasound equipment capable of the higher frequencies to observe and model the behaviour of acoustic speckle. In this way, a more accurate, frequency-dependent model may be developed for acoustic shadow detection.

4. A clearer definition of acoustic shadowing is required as an acoustic shadow exhibits a transition region before presenting as a dark area of signal loss.
Shadow detection algorithms require quantitative definitions in order to accurately determine the beginning of an acoustic shadow. This limitation results in inconsistencies between shadow detection methods as performance varies depending on the definition of the boundaries of an acoustic shadow. Future work to address this involves modeling the transition region to quantitatively define when an acoustic shadow begins.

Limitations of automated placenta segmentation are that:

1. There are noticeable areas of the placenta that exhibit the most frequent errors, namely the boundary between the uterine wall and the placenta. This results in segmentation error that usually falsely identify the uterine wall as part of the placenta. This may be due to the different texture of the placenta at the boundaries. As the body of the placenta occupies the majority of the placental area, there is relatively limited image data for an algorithm to train in recognizing placental boundary texture. Future work should include more data to train boundary areas or the addition of an attention network to specify that boundary regions are classified differently than the placenta body. This can lead to the segmentation of more anatomical detail beyond simply labelling placenta and non-placenta region. A full fetal anatomy classifier may be developed in the future using similar methods to identify regions such as the placenta cord insertion site and the umbilical cord. This requires further ground truth labeling and data collection of sufficient images with the anatomy of interest visible.

2. The majority of the data used was selected to exhibit placenta visible for an annotator. Thus, there was limited training on images where a placenta is difficult to see visually. In addition, all ultrasound images were taken by expert sonographers. Thus, the segmentation method was demonstrated only on images where the ultrasound setting were set specifically for fetal scans. Although there is high usability in the method in that no expert configuration of the segmentation algorithm is required, there is still a barrier of expertise for configuration of the base ultrasound machine. The segmentation method has
not been tested on poorly captured ultrasound data by non-experts. This is a limitation in demonstrating the usability of the method for novice users who may not know how to optimize the operation parameters of an ultrasound machine. To increase the versatility of segmentation, future work may include training on highly obscured placenta or low quality ultrasound images to develop a segmentation model that can outperform manual analysis and be usable by novices.

Limitations binary placenta classification are that:

1. There are limitations with the classifier that attempts to predict the clinical outcome of a placenta as with disease or without disease. The poor performance of the classifier indicates that there are limitations in the method and likely the dataset used. As the dataset consisted of still frame images from placental sweeps, all images of a sweep are labeled with the same outcome in reference to the patient clinical charts. Thus, images containing healthy regions of the placenta in the sweep would also be labeled with an adverse outcome. This decreases the ability for the CNN to train effectively. Future work would include improving the CNN design to handle placenta sweep data by stitching images together or introducing spatial information such that all sweep frames are considered part of the same placental volume. This requires much more data acquisition to provide enough patient sweep sets to train the network.

2. The classification was tested on a limited range of architectures and network parameters. In deep learning, there is a common limitation of abstracting the mathematical relationships between the parameters, kernels, and layer placement. This is due to the complexity of the network and understanding the best design of a neural network is still an active field of research [10]. Future work to improve placenta classification methods include exploring different architectures and iterating through the architecture parameters to determine the best method to organize the layers of a neural network to identifying features of a placenta ultrasound image.
3. The study conducted for placenta classification utilized only ultrasound image data. Numerical patient clinical data, such as demographics and risk factors, may also assist in classification of placenta ultrasound images. The limitation in the study conducted was the patient clinical data only contained records for 364 patients, which is too limited to train a neural network with. Future work includes the acquisition of a large patient dataset, with both ultrasound and numerical clinical data. The ultrasound and numerical data can then be potentially combined into a neural network such that a network can associate both ultrasound image features and important risk factor threshold with placenta disease.

This thesis presents a system pipeline for analyzing placenta ultrasound images that may aid physicians in assessing the health of the placenta. The results of placenta classification of whether a placenta image appearance is normal or abnormal indicates that the method has the potential to address several challenges. First, the method is capable of achieving high classification accuracy, indicating that it may be usable to aid screening of placental health. Second, the method is automated after receiving a single configuration parameter for acoustic shadow detection and can perform placenta classification without expert intervention. This allows for the method to potentially be usable in areas lacking specialized maternal-fetal medicine physicians for earlier and frequent screening of placental health. Third, the method allows for classification and visualization of placenta features that are associated with an abnormality. This allows potential early study of the placenta through ultrasound images and the identification of regions of interest for localized pathological study. The results and limitations motivate future studies to improve placenta analysis to eventually improve the early detection of placental disease.
Bibliography


89


