Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Nasim Katebijahromi Date

Detection of Adverse Events in Pregnancy Using a Low-Cost 1D Doppler Ultrasound Signal

By

Nasim Katebijahromi
Doctor of Philosophy

Computer Science and Informatics

Gari D. Clifford, DPhil.
Advisor

Reza Sameni, Ph.D.
Co-advisor

Ali Bahrami Rad, Ph.D.
Committee Member

Qiao Li, Ph.D.
Committee Member

Accepted:

Dean of the James T. Laney School of Graduate Studies

Date

Detection of Adverse Events in Pregnancy Using a Low-Cost 1D Doppler Ultrasound Signal

By

Nasim Katebijahromi
B.Sc., Shiraz University, Iran, 2013
M.Sc., Shiraz University, Iran, 2016
M.Sc., Emory University, GA, 2020

Advisor: Gari D. Clifford, DPhil.
Co-advisor: Reza Sameni, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2021

Abstract

Detection of Adverse Events in Pregnancy Using a Low-Cost 1D Doppler Ultrasound Signal
By Nasim Katebijahromi

Fetal maternal mortality is an enormous global health challenge, affecting over 2.6 million families annually. The burden is most heavily felt in low-and middle-income countries (LMICs) due to systemic healthcare issues related to inequity, limited funding for medical technology, and poor infrastructure for delivering and maintaining technology. Fetal growth restriction is increasingly recognized as an important contributor to fetal health problems in LMICs. One of the effective approaches to detect fetal developmental issues is tracking fetal heart rate variability (FHRV). FHRV is also an indicator of fetal brain development since it is influenced by the autonomic nervous system, which evolves during pregnancy. Therefor, accurate estimation of gestational age using FHRV patterns could help to detect fetal brain developmental issues and potential cases of small for gestational age. This thesis aims to enhance fetal health monitoring for disadvantaged populations by developing AI-enabled edge computing devices which are intuitive to use even for low-literacy populations. Specifically, this work presented machine learning methods to analyze one-dimension fetal Doppler ultrasound signals (1D-DUS), which have been collected using a low-cost mobile health system.

Developing accurate models to capture the underlying dynamic of 1D-DUS is a challenging task. Since 1D-DUS is nonstationary, highly susceptible to noise and movement, and has a transient nature. Using additional device for recording another data modality or labels of the beat intervals can mitigate the effect of highly variable morphology. However, these solutions significantly complicate the use and raise the cost of the smartphone-mediated perinatal screening system. Therefore, this thesis aimed to underline the importance of taking challenges in LMICs into account for developing accurate machine learning methods and addressed the challenges via two approaches: 1) developing an unsupervised probabilistic segmentation method to estimate FHRV metrics from 1D-DUS recordings. 2) Developing a deep sequence learning model with an attention mechanism for automatic feature extraction and estimation of fetal gestational age. This model is the first attempt to estimate gestational age from only Doppler signals and outperforms previous attempts based on multiple signals. The developed methods will ultimately run on-device and interact with the healthcare workers or mothers directly. This work could assist traditional birth attendants in rural areas with a decision support system to identify patients with possible pregnancy-related abnormalities for early triage and intervention.

Detection of Adverse Events in Pregnancy Using a Low-Cost 1D Doppler Ultrasound Signal

By

Nasim Katebijahromi
B.Sc., Shiraz University, Iran, 2013
M.Sc., Shiraz University, Iran, 2016
M.Sc., Emory University, GA, 2020

Advisor: Gari D. Clifford, DPhil.
Co-advisor: Reza Sameni, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2021

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my research advisor Prof. Gari Clifford for giving me a chance to be part of his research group. Words are certainly not enough to express my gratitude for his endless support, expert guidance, and encouragement. It is always a blessing to chat with him. He is super smart, and his thoughtful comments helped me a lot to direct my research path. I would like to thank Prof. Reza Sameni for all his efforts and suggestions. He was always available and patiently helped me through my research.

This thesis would not have been possible without the support of many. My thanks go to Peter Rohloff, Boris Martinez, Enma Coyote Ixen, Michel Juarez, Rachel Hall-Clifford, Faezeh Marzbanrad, Maya Health Alliance, and the lay midwives of Tecpan, Chimaltenengo. Special thanks to Lisa Stroux and Camilo Valderrama. I owe a great deal of achievement in my research to their amazing work. I am very grateful for having my friends at Clifford Lab, who made this journey wonderful. Thank you for all the good memories! I have been lucky to have my friends, Parisa and Mohammadreza, who have always been there for me and supported me through all the ups and downs in my life. Finally, I would like to thank my family for being so supportive. I am deeply grateful to my husband Yashar for his endless love, continuous support, and companionship throughout our Ph.D. journeys. He is the one who kept me going and made me smile through the difficult times of research and life. Last but not least, I am thankful for my loving parents and brother for their endless support, patience, and encouragement throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis contributes to approaches for enhancing healthcare in disadvantaged populations by developing AI-enabled edge computing devices that are intuitive to use, even for low-literacy populations. In particular, this work focuses on improving fetal-maternal monitoring for rural pregnant disparity populations in highland Guatemala, where one-dimensional Doppler ultrasound (1D-DUS) recordings have been collected over the last five years using a low-cost device [237]. There are a lot of studies on fetal cardiac signals in the field, however most of them are not suitable for low-resource settings. This thesis proposes methods based on issues identified during pilot research and a randomized control trial [166, 165, 237, 234, 237]. The developed algorithms will ultimately run on-device and interact with the healthcare workers or mothers directly.

1.1 Motivation

Fetal maternal mortality is an enormous global health challenge, affecting over 2.6 million families annually [38, 148]. The burden is most heavily felt in low-middle income countries (LMICs) where 98% of total perinatal deaths occur [284]. Among LMICs, Guatemala suffers one of the highest rates of perinatal morbidity and mor-

tality in Latin America which is reported as 73 deaths per 100,000 live births. This rate can more than double among rural indigenous women, particularly the Mayan population. [243]. Alarmingly, major causes of fetal maternal deaths in low resource regions are preventable by addressing inequalities and discrimination in healthcare system that affect maternity care. There are major factors that perpetuate the high perinatal mortality rates in these areas such as chronic health problems, a lack of knowledge about obstetric danger signs and a lack of access to professional medical care and routine perinatal screening due to economic status, language, and culture. Therefore, home delivery with the assistance of traditional birth attendants (TBAs) is common and lack of specialized training of TBAs leads to ineffective detection of fetal maternal compromise [122, 165].

In LMICs, fetal developmental issues, specifically, intrauterine growth restriction (IUGR) is known to be the most prevalent fetal complication [151]. In high-income countries, obstetricians detect suspicious cases of IUGR by Doppler ultrasound imaging [154]. However, the cost of purchase, the technical skills required for maintenance and the user-dependent accuracy have limited the application of this technique in resource-limited settings [189]. Therefore, the need for low cost and accurate monitoring and diagnostics system is acute, particularly in low resource regions of the world.

One of the strategies identified as being crucial for the detection of fetal health problems and as an indicator for fetal development is cardiac assessment [196]. As previous studies have reported, fetal heart rate variability (FHRV) is associated with development of fetal autonomic nervous system (ANS) and helps to detect IUGR [238, 110, 240, 248, 171, 170].

An affordable fetal monitoring system was introduced by Stroux *et al.* [237] to assist TBAs in rural areas of Guatemala. The mhealth monitoring system consists of a low-cost 1D Doppler transducer connected to a smartphone and a standard

oscillometric blood pressure monitor. The app guides the user to record the data by helping to find the fetus and to record fetal cardiac activity. The app also guides the user through a checklist - the Pregnancy Risk Assessment Monitoring System (PRAMS) [48]. This is presented to the TBA through culturally appropriate pictures and audio prompts in the local language to help to identify concerning signs and symptoms during pregnancy. User responses are combined with the results of the blood pressure readings and Doppler recording analysis to identify concerning signs and symptoms during pregnancy and then provide an alert to a healthcare worker for decisions on appropriate interventions.

Doppler transducer has the capability of capturing the flow of blood through the heart's chambers and valves which conveys important information regarding the hemodynamic status and cardiovascular adaptation of a fetus in the face of several perinatal complications. Previous studies used signal processing techniques to estimate beat-to-beat variability and extract the characteristics of the 1D-DUS signal. Hence, the location of the beats was determined by hand labeling or providing simultaneous recordings. While additional recording technique enhances the analysis, it is not suitable for the application of interest due to increasing the complication and raising the cost of the screening system. Furthermore, manual identification of the cardiac valve timings is time-consuming, requires special expertise, and is subject to visual errors. Furthermore, 1D-DUS signal is nonstationary, highly susceptible to noise, and has a transient nature which complicates the extraction of the information. As such, improved approaches for automatic feature extraction and modeling were investigated in this research.

1.2 Aim of this thesis

This thesis aims to provide methods for automatic fetal 1D-DUS feature extraction and detection of adverse events in pregnancy. To this end, signal processing and machine learning techniques were used to learn the underlying dynamics in fetal cardiac activity. To achieve our final aim, the following novel research was performed:

- A method to segment the 1D-DUS signal in an unsupervised manner and extract fetal heart rate variability metrics enabling by hidden semi-Markov model. The proposed probabilistic segmentation method employs envelope and spectral features as input of data transformation block which includes time-delay embedding and kernel density estimation to reduce the variability while keeping useful information in state transitions. The beat onsets and fetal cardiac beat-to-beat intervals were estimated from the segmentation results. Comparison of heart rate variability metrics estimated from 1D-DUS and fetal electrocardiogram shows the effectiveness of the presented model.

- A method to estimate gestational age using deep sequence learning model consists of convolutional Long short-term memory networks. Time-frequency features were extracted from Doppler signals and regularized before feeding to the network. The presented work is the first attempt to estimate gestational age from only Doppler signals, and outperforms previous attempts based on multiple signals.

- A method to detect noisy gestational age labels recorded in Guatemala site based on multiple observation fusion. The proposed label evaluation metric gives insight to improve the fetal monitoring system and helps to improve the analysis by using training strategies to learn from noisy labels.

- A method for learning hierarchical relationship in cardiac signal to estimate ges-

tational age. This model incorporates the information in long-term and short-term fetal cardiac activity. The training strategies to increase the generalization on less frequent labels helped to improve the results.

1.3 Thesis outline

This thesis comprises seven chapters besides this introduction, all of which (except for the conclusion) have been published or are under review in key journals and conferences in the field (see section 1.4).

Chapter 2 presents the background of this thesis. The chapter reviews the fetal cardiac monitoring approaches in LMICs. Then, the chapter describes the limitations and challenges of using medical technologies in low-resource settings. The chapter concludes by introducing low-cost mobile health technologies to overcome LMICs barriers.

Chapter 3 provides the details of data collection and labeling process. Then, the chapter introduces the GUI used for annotating the quality and beats of the fetal cardiac signals.

Chapter 4 presents the unsupervised segmentation method to estimate fetal heart rate variability. This chapter also introduces data transformation method based on time-delay embedding to reduce the effect of signal variability in the segmentation part of the model.

Chapter 5 proposes a deep learning approach for automatic extraction of discriminative information from 1D-DUS signals. This chapter presented an end-to-end training of the network to map the generated time-frequency features to the gestational age labels.

Chapter 6 provides a method to evaluate the quality of the recorded gestational age labels. Then this chapter provides the analysis of the defined error based on

variables of the study.

Chapter 7 extends the method introduced in Chapter 5 by presenting an approach to estimate gestational age using the deep hierarchical attention network.

Finally, Chapter 8 presents a summary of contributions, limitations, and possible future work.

1.4 List of publication

Work in this thesis has been published in the following journals and conference:

- Valderrama C. E., Katebi, N., Marzbanrad, F., Rohloff, P., Clifford G. D. A review of fetal monitoring for well-being assessment, with a focus on low-and middle-income countries. Physiological Measurement. 2020 Oct 26.
  (This publication appears in its entirety in Chapter 2)

- Valderrama C. E., Marzbanrad F., Stroux L., Martinez B., Hall-Clifford R., Liu C., Katebi N., Rohloff P., Clifford G. D. Improving the quality of point of care diagnostics with real-time machine learning in low literacy LMIC settings. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies 2018 Jun 20 (pp. 1-11).
  (This publication appears in Chapter 3).

- Valderrama C. E., Stroux L., Katebi N., Paljug E., Hall-Clifford R., Rohloff P., Marzbanrad F., Clifford G. D. An open source autocorrelation-based method for fetal heart rate estimation from one-dimensional Doppler ultrasound. Physiological Measurement. 2019 Feb 25;40(2):025005. (This publication appears in Chapter 3).

- Katebi N., Marzbanrad F., Stroux L., Valderrama C. E., Clifford G. D. Unsupervised hidden semi-Markov model for automatic beat onset detection in

1D Doppler ultrasound. Physiological Measurement. 2020 Sep 4;41(8):085007. (This publication appears in its entirety in Chapter 4).

- Katebi N., Sameni R., Clifford G. D. Deep Sequence Learning for Accurate Gestational Age Estimation from a $25 Doppler Device. ML4H 2020, NeurIPS 2020, December 11, 2020. (This publication appears in its entirety in Chapter 5).

- Katebi N., Sameni R., Clifford G. D. Detection and relabeling of noisy gestational age recordings. (In submission to Physiological Measurment) (This publication appears in Chapter 6).

- Katebi N., Sameni R., Clifford G. D. Hierarchical attention network for gestational age estimation. (In submission to IEEE Transactions on Biomedical Engineering) (This publication appears in Chapter 7).

Chapter 2

A review of fetal cardiac monitoring, with a focus on low-and middle-income countries [1]

2.1 Abstract

There is limited evidence regarding the utility of fetal monitoring during pregnancy, particularly during labor and delivery. High income countries rely on consensus 'best practices' of obstetrics and gynecology professional societies to guide their protocols and policies. Protocols are often driven by the desire to be as safe as possible and avoid litigation, regardless of the cost of downstream treatment. In high-resource settings, there may be a justification for this approach. In low-resource settings, in particular, interventions can be costly and lead to adverse outcomes in subsequent pregnancies. Therefore, it is essential to consider the evidence and cost of different fetal monitoring approaches, particularly in the context of treatment and care in low-to-middle income countries.

This chapter includes the standard methods used for fetal monitoring, with particular emphasis on fetal cardiac assessment which is a reliable indicator of fetal well-being. An overview of fetal monitoring practices in low-to-middle income counties, including perinatal care access challenges, is also presented. Finally, an overview of how mobile technology may help reduce barriers to perinatal care access in low-resource settings is provided.

2.2 Introduction

Perinatal complications account for 40% of the perinatal and maternal deaths worldwide [272]. Low-and middle-income countries (LMICs) contribute approximately 90% of total births, and 98% of the total perinatal deaths [268, 213, 38, 262].

The perinatal mortality rate is defined as the sum of the number of stillbirths and deaths occurring during the first seven days of life, per 1000 live births. In 2018, this rate stood at 19 per 1000 in LMICs, whereas in upper-middle and high-income countries, there was an average of seven and three deaths per 1000 live births, respectively [244]. The highest perinatal mortality rates have been reported for countries in Sub-Saharan Africa and South-Asia (28% and 26%, respectively) [244] and may be underreported [155, 195]. At the beginning of the twentieth century, the perinatal mortality rate in high-income countries (HIC) was as alarmingly high as it currently is in LMICs, but was effectively reduced by the expansion of antenatal care coverage, extended indications for Cesarean sections, and the introduction of perinatal screening technologies (cardiotocography (CTG), ultrasound, amnioscopy, amniocentesis, and pH-meter) [69, 149, 78, 93].

The most common causes of perinatal deaths are preterm birth-related complications (35%), intrapartum-related events (24%), and sepsis (15%) [243]. Studies conducted in LMICs have reported significant issues with prematurity, birth asphyxia,

maternal hypertensive disorders, and septicemia being the most common causes of perinatal death [10, 157]. Fetuses and newborns are also disproportionately affected by infections, including syphilis, malaria, and animal and vector-borne diseases, leading to elevated mortality and morbidity [101, 91].

Asphyxia, one of the most common causes of death during childbirth [90, 149, 260, 257], involves oxygen deprivation arising from obstruction of the placental blood flow, which may be rooted in maternal pre-eclampsia, placental abruption, or umbilical cord accident. The high death rate associated with asphyxia is mainly due to poor delivery management. Signs of asphyxia can be identified via fetal heart rate monitoring [76], and timely detection and intervention can reduce the risk of irreversible organ damage and identify cases requiring rapid deliveries [91]. However, this basic monitoring procedure is not often practiced in LMICs.

Low birth weight (LBW), common among preterm (<37 weeks) or small-for-gestational-age (SGA) babies, is documented in 70-80% of the perinatal deaths [146, 10, 157]. In LMICs, approximately 60% of LBW newborns are SGA [151], which, in these countries, is often ascribed to intrauterine growth restriction (IUGR) [63, 151]. IUGR can develop as a consequence of maternal vascular problems, malnutrition, or placental malfunction [282].

While fetal cardiac assessment has been in use over the past four decades to diagnose, monitor, or predict adverse fetal conditions throughout pregnancy [154, 273], there is still insufficient evidence with regards to its contribution to improved perinatal outcomes [221]. As a result, the World Health Organization (WHO) does not currently recommend continuous cardiotocography during labor for assessment of fetal well-being in healthy pregnant women undergoing spontaneous labor, but rather a periodic point of care auscultation [275].

To explore the potential future directions for fetal monitoring in low-resource settings, this review presents an insight into fetal cardiac assessment, briefly explaining

the affordability and applicability to each stage of pregnancy. Finally, we provide an overview of how mobile technology may reduce barriers to access perinatal care in poor-resource settings.

2.3 Fetal cardiac circulation

The human heart develops within the first six weeks of gestation [88], with the majority of its functionality achieved by the eighth week of gestation [19]. The development begins with a primary heart tube, which evolves into the four-chambered adult heart structure, composed of two atria and two ventricles. Fetal circulation is unique in that blood is oxygenated in the placenta rather than in the lungs [231].

The oxygenated blood from the placenta (umbilical vein) flows into the fetal liver and later, the inferior vena cava via the ductus venosus. On the one hand, the majority of oxygenated blood flows directly from the right atrium to the left atrium through the foramen ovale, and subsequently to the left ventricular to be pumped to the aorta [231, 82]. On the other hand, the remaining oxygenated blood passes from the right atrium to the right ventricle and subsequently to the pulmonary vein. As fetal lungs are non-functional, a significant percentage of the blood in the pulmonary vein passes into the aorta via the ductus arteriosus [231]. The blood sent to the aorta circulates to the fetal brain and tissues. Finally, deoxygenated blood is transported to the placenta via two umbilical arteries [82]. After birth, the foramen ovale closes, resulting in occlusion of the ductus venosus and arteriosus, and to the separation of the pulmonary and circulatory functionalities [231].

2.3.1 Control of fetal heart rate

The fetal heart rate (FHR) represents the reciprocal of the interval between two successive fetal heartbeats. The heartbeats are controlled by cardiac muscle cells

located in the myocardium [231]. The cardiac cells are categorized as myocardial contractile cells and myocardial conductive cells. The contractile cells stimulate the contractions required to pump blood throughout the body, whereas the conductive cells are the autorhythmic cells responsible for the heart's electrical activity.

The majority of the conducting cells are located in the sinoatrial node (SAN), also called the pacemaker. The SAN initiates action potentials, resulting in the contraction of the atria at the onset of systole [231]. The action potential is propagated via the atrioventricular (AV) node to the bundle branches and Purkinje fibers located within the ventricular walls. This impulse initiates ventricular contractility, which, in turn, pumps blood to the pulmonary veins and the aorta, to mark the end of systole. The impulse then leaves the ventricles, marking the onset of diastole, during which the ventricular walls are repolarized. The electrical and mechanical events of a heart contraction generate the cardiac cycle, which is measured as the number of beats per minute (bpm).

Throughout pregnancy, the pace of fetal cardiac activity is controlled by the autonomic nervous system (ANS), baroreceptors, and chemoreceptors [82]. The ANS is comprised of the sympathetic and parasympathetic nervous systems. The sympathetic system accelerates the heart's electrical activity, yielding a faster FHR. The parasympathetic system, on the other hand, has the opposite effect on the FHR.

The balance between the sympathetic and parasympathetic nervous systems sets the baseline of the heart rate. However, as the sympathetic system matures earlier than the parasympathetic system, the FHR is higher in the first months of gestation. At 15 weeks gestation, the average FHR is 60 bpm. With the advancement of pregnancy, and the evolvement of the parasympathetic system, the FHR increases to approximately 110-160 bpm [197].

Figure 2.1 illustrates how the sympathetic and parasympathetic systems affect the FHR across gestation, as reported by [258]. On analysis of traces of 61 healthy

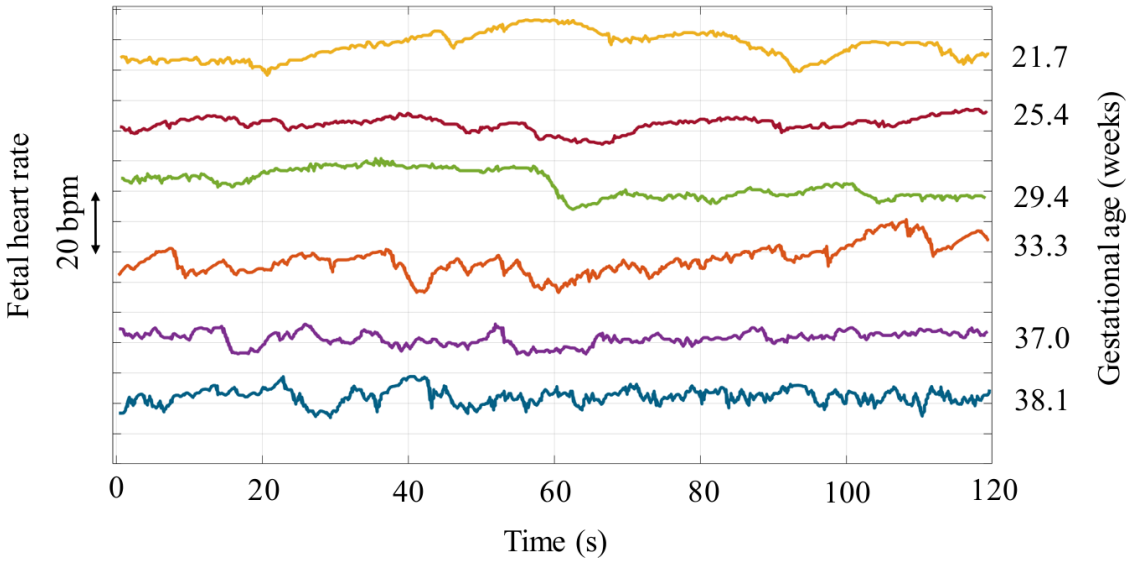


Figure 2.1: Illustration of the variation in FHR during gestation weeks 22-38. Note that the vertical axis has an arbitrary offset. FHR drops by about 15 bpm from week 25 to week 40 of gestation [128]. Adapted from [258].

pregnant women, a short-term increase in FHR variability during the last trimester was noted. In contrast, long-term variability in FHR was most pronounced during the early gestational period.

The other two mechanisms that regulate the fetal heart rate are the baroreceptors and chemoreceptors. Barocepters are located in the aortic arch, carotid arteries, and brain stem. When blood pressure increases, baroreceptors signal the vagal nerve to slow down the heart rate, which then reduces blood pressure. In response to the blood pressure decrease, baroreceptors reduce the parasympathetic tone and stimulate an increase in the fetal heart rate and blood pressure.

The chemoreceptors, found in the aorta, carotid artery, and brain stem, impact the fetal heart rate via its oxygen level-sensing capacities. When the oxygen level decreases, the FHR is accelerated to increase the oxygen input rate from the placenta. However, when the oxygen level reduction is abrupt (hypoxemia), the chemoreceptors trigger a vagal response, resulting in a reduction in heart rate and an increment in blood pressure.

2.4 Fetal heart monitoring techniques

Fetal heart monitoring technologies can be categorized as either intermittent auscultation (IA) or electronic fetal monitoring (EFM) methods. IA techniques focus on verifying fetal cardiac performance by counting the number of beats over short periods, most commonly measured with a Pinard fetoscope, DeLee fetoscope, or handheld Doppler device [39]. EFM methods identify fetal stress or distress based on FHR variability, commonly performed via cardiotocography (CTG) [177]. It provides continuous information on the FHR, for a period of 10-60 minutes, using autocorrelation to obtain the average FHR over a specific window, which is generally every 3.75 s [168].

EFM techniques can be categorized into invasive and non-invasive methods. In the invasive mode, the fetal electrocardiogram (fECG) is taken directly from the fetal scalp [25]. Although the invasive technique is more accurate than non-invasive modalities, its use is limited to the intrapartum period, when the membranes are ruptured. In contrast, non-invasive methods are only employed during the antenatal period. Non-invasive methods extensively described in the literature include CTG, abdominal fECG, phonocardiography (fPCG), and fetal magnetocardiography (fMCG) [84, 207, 267, 23, 2].

2.4.1 Fetal phonocardiogram

The fetal phonocardiogram (fPCG) is an electronic extension of the Pinard and DeLee stethoscopes. Similar to the stethoscope, fPCG is an IA technique in which a microphone is placed on the maternal abdomen to listen to fetal heart sounds [207]. The audible heart sounds correspond to the closure of the fetal valves during the cardiac cycle [139]. The closure of the mitral and tricuspid valves generates a sound called S1, and the closure of the semilunar valves (pulmonary and aorta) generates a

sound called S2. Both S1 and S2 have low acoustic energy and are affected by noises such as environmental noise, as well as other maternal and fetal physiological sounds, such as breathing, fetal movements, and maternal circulation [253]. The technique can extract cardiac timing and intensity of fetal heart sounds, which can carry useful diagnostic information [2].

The fPCG can be used during the antepartum phase (gestational week $\geq$ 24) [253]. Although fPCG is an alternative to the traditional ultrasound used in perinatal management [100], it is underutilized [2] and suffers from significant challenges related to signal acquisition and processing. Further research is needed to improve fPCG to compete with standard fetal monitoring methods, i.e., CTG and ultrasound imaging.

2.4.2 One-dimensional Doppler ultrasound

One-dimensional Doppler ultrasound (1D-DUS) estimates FHR by measuring the Doppler shift between ultrasound beams transmitted and received from the mechanical heart movements and blood flow. The Doppler magnitude frequency shift fD, is described as [136]:

$$f_D = \frac{2f_o}{c} V \cos\theta, \tag{2.1}$$

where f_D is the measured change in frequency (Hz), f_o the frequency of emitted ultrasound transducer in Hz, c the speed of sound in soft tissue in m/s, V the velocity of the reflecting interface in m/s and θ is the angle between the ultrasound beam and the surface in radians.

The transmitted beam travels across various anatomical structures, from the skin surface, through the maternal skin and subcutaneous tissue, and then finally reaches the uterine muscles, the amniotic sac, and the fetal heart [163]. The fetal heart movement reflects the ultrasound beam, and propagates the ultrasound waves in the reverse order. The distance between the DUS transducer and fetal heart depends

on the maternal phenotype, which varies among nationalities [269], socioeconomic status, as well as body mass index [179].

The shifted Doppler frequency is usually demodulated via the phase-quadrature demodulation, in which the received signals are mixed with the carrier signals $\sin 2\pi f_o t$ and $\cos 2\pi f_o t$ [72]. The demodulated signal is then autocorrelated to estimate the cycle period of the heartbeat rhythm [219].

Doppler ultrasound includes two different modes - the continuous wave (CW) and the pulsed-wave (PW). In CW, two piezoelectric crystals continuously monitor the reflection of the emitted wave. In PW, one piezoelectric crystal alternates between sending and receiving the sound waves. The dual functionality of CW Doppler enables the measurement of higher velocities. However, as velocities are measured in the same line of interrogation, it is impossible to know the origin of the velocity. In contrast, PW measures slower velocities, but the emitted sound waves are associated with the received waves, thus enabling the detection of the structure's distance reflecting the wave.

CW is primarily integrated in hand-held Doppler transducers, while PW is used in standard CTG machines. Hand-held Doppler transducers are used during the intrapartum and antepartum periods after the 20th gestational week, to measure heart rate variability metrics as an indication of fetal wellbeing.

[158] carried out a randomized, controlled trial that showed that fetal monitoring using hand-held DUS transducers could detect a similar number of prolonged or late decelerations as ultrasonography. Additionally, the hand-held Doppler devices detected a substantially larger number of late and prolonged decelerations than the Pinard stethoscope. The potential of hand-held Doppler devices has also been reported by [66], following a randomized controlled study that demonstrated that they provide the same level of safety for screening and monitoring as cardiotocography in low-risk pregnancies.

2.4.3 Cardiotocography

Cardiotocography (CTG) is the simultaneous and continuous measurement of FHR and uterine pressure, often detected as uterine contractions (UC), and is a standard method for assessment of fetal wellbeing [98]. To record the FHR, the medical assistant applies a gel on the maternal abdomen and the ultrasound transducer. The transducer is moved across the maternal abdomen while the technician listens for an audible version of the Doppler signal, in an attempt to identify the spot with maximum fetal heart rate impulse (as opposed to maternal arterial flow) [82].

CTG is used for fetal monitoring starting at the 20th week of gestation [66], but most commonly indicated after the 28th week of gestation [216]. Although CTG is widely used, it suffers from high intra- and inter-interpreter variability [241, 36, 40, 37, 112], resulting in low specificity. To reduce this subjectivity, [61] introduced a computerized version of the CTG. A Cochrane review of two studies (469 subjects) concluded that the mortality rate in a population monitored by computerized CTG was four times lower than in the population monitored by visual CTG (0.9% vs. 4.2%) [98].

Computerized CTG has also been used in recent years to develop artificial intelligence methods to detect abnormal FHR patterns, achieving comparable results to clinical assessment of the CTG [50, 265, 86]. Notably, these artificial intelligence-based CTG systems have shown the potential to discriminate between normal and IUGR fetuses [238, 222].

Although CTG is a standard method used for fetal monitoring in high-income countries, controlled clinical trials have not provided evidence of its benefits; CTG was associate with a 20% increase in Cesarean interventions with no improvement in fetal outcomes [66]. In addition, the use of CTG was not associated with statistically significant improvements in perinatal outcomes as compared to traditional intermittent auscultation methods [127, 176].

2.4.3.1 The nonstress test

The nonstress test (NST) monitors FHR patterns for at least 20 minutes, and is designed to identify accelerations associated with fetal movements [160]. This test calculates the baseline FHR, which is later used to measure long-term and short-term variability, episodes of high and low variation, acceleration, and deceleration. Test results are considered normal (reactive) when more than two accelerations occur within 20 minutes of observation. In contrast, a non-reactive result is when, no more than one acceleration occurs within 40 minutes [13]. The NST has a low false-negative (0.3%), but a high false-positive (50%) rate [71]. Of note, the test carries no risk of inducing any uterine contractions.

2.4.3.2 Contraction stress test

The contraction stress test (CST) is based on the premise that contractions, induced using oxytocin or nipple stimulation [160], trigger a hypoxic state [221]. A healthy fetus can tolerate this hypoxic state, whereas a non-healthy fetus will respond with late FHR decelerations [160]. Although this method has a low false-negative rate (0.04%) [81] and a lower false-positive rate than NST (30% vs. 50%) [13], it requires an intravenous intervention, which increases the risk of fetal hypoxia and of induction of preterm birth [160].

2.4.3.3 Acoustic stimulation

Acoustic stimulation is a variation of NST, in which if there are no fetal movements, vibroacoustic stimulation (usually with a laryngeal stimulator) may be activated for 3 seconds on the maternal abdomen over the fetal head. This is performed to 'awaken a sleeping fetus', before initiation of the NST [160]. The artificial larynx produces a vibratory stimulus of 80 Hz that causes a healthy fetus to increase physical activity, as measured by an increase in FHR. Its advantages include shortening of the NST by

10 minutes [51], and reduction of the number of non-reactive states, without affecting readability [223, 224]. In cases of a non-reactive result, the acoustic stimulation is repeated for five minutes. If the test is still non-reactive, a fetal biophysical profile or CST is indicated [160].

2.4.4 Fetal electrocardiogram

The fECG records the complex electrical activity of the fetal heart. The main components of the ECG signals are P, Q, R, S, and T waves. The P wave represents atrial depolarization, which is followed by the atrial contraction (atrial systole). The atrial contraction is extended to the QRS complex, which corresponds with ventricular depolarization, with ventricles contracting at the peak (R wave). The ventricular contraction lasts until the ST-T wave, which corresponds with ventricular repolarization and relaxation. fECG can be captured in an invasive manner during the intrapartum period, when the cervix is dilated, and the fetus scalp is visible, or in a non-invasive manner, starting from the second trimester. The fECG is also used to complement CTG at intrapartum to reduce unnecessary Cesarean sections [256, 70].

2.4.4.1 Invasive fetal electrocardiogram

The invasive fetal electrocardiogram (invasive- fECG) requires the rupture of the membranes to introduce electrodes, via the cervix, and to place them on the fetal scalp [105, 106, 107]. This technique processes the recorded signals to visualize the P and T waves, as well as the QRS complexes.

Scalp fECG has been used as a complementary technique during intrapartum FHR monitoring [12, 187]. The morphology of the ST segment is analyzed to find patterns associated with uterine complications [145, 144]. Invasive fetal ST can be captured and analyzed from the 36th gestational week and is indicated in high-risk pregnancies when a non-reactive CTG is obtained, or labor is induced by oxytocin. Although its

use has shown to effectively reduce neonatal encephalopathy [186, 187], randomized control trials of this technology have yet to demonstrate a clear benefit.

2.4.4.2 Non-invasive fetal electrocardiogram

Non-invasive fECG measures the electrical activity of the fetal heart via electrodes which are placed on the maternal abdomen [142]. This technique is indicated from the 18th week of gestation [207], and therefore has much wider applicability than invasive ST analysis and can replace Doppler auscultation for the fetal heartbeat.

Although abdominal fECG signals have a relatively low amplitude (microvolts), it can provide a more accurate estimate of beat location when compared to the CTG, and hence a more accurate quantification of fetal heart rate variability indexes [119, 54, 121]. The morphology and beat-to-beat heart rate variability estimated from fECG are established indicators of pre-eclampsia and IUGR. For instance, a study conducted on 106 patients (30 healthy, 44 mild pre-eclampsia, and 32 severe pre-eclampsia subjects) at 34-40 weeks of gestation, reported that FHR variability indices were associated with the suppression of fetal biophysical activity and the development of fetal distress in women suffering from severe pre-eclampsia [143]. Similarly, [254] assessed the impact of IUGR on FHR variability indices extracted from abdominal fECG recordings of 20 control and 15 IUGR singleton pregnant women. While the authors identified clear P-QRS-T complexes in all cases, prolonged QT intervals were measured in IUGR fetuses.

Over the last 30 years, a variety of methods have been proposed for extracting and processing fECG signals [207, 116]. Methods range from adaptive filtering [193, 220, 164] to non-adaptive approaches such as, independent component analysis [209], principal component analysis [4], wavelet transforms [47, 279] and neural networks [20, 14, 32]. Many of these techniques suffer from significant limitations due to causality and signal stability. Other approaches based on generalized eigenvalue decomposition

have shown more promise [211]. In [53, 31], it was shown that this approach was able to accurately resolve both QT interval and ST elevation/depression from non-invasive fECG. However, this promising result has yet to be applied in a randomized clinical trial to demonstrate efficacy.

2.4.5 Fetal magnetocardiography

Fetal magnetocardiography (fMCG) uses a sensitive, superconducting sensor to measure the magnetic field of fetal heart activity [129, 207, 133]. The fMCG provides a waveform almost identical to that of fECG, but at a higher signal-to-noise ratio, and with a higher resultant quality of waveform [196, 133]. The higher quality enables the classification of arrhythmias and detection of congenital heart diseases [124], as well as the ability to assess fetal neurological development [258].

The fMCG is used from the 20th gestational week [196]. Yet, although it provides good-quality waveforms, it is not routinely used in perinatal care due to its higher costs, i.e., the need for a shielded room, and highly skilled personnel [196, 133]. Alternative methods, such as the abdominal fECG or the hand-held Doppler, can be used at any time during pregnancy, and can even be performed at home by the patients themselves [207].

2.5 Ultrasound imaging

Ultrasound imaging is considered the gold standard for fetal monitoring in high-income countries [154, 274]. It evaluates fetal growth, fetal cardiac structure and function, and fetal, uterine, and placental blood circulation. Ultrasound imaging is usually indicated in the second trimester, particularly after the 20th week of gestation, with a scan recommended by the WHO before week 24 [266, 274, 273].

Ultrasound imaging is known to effectively assess pregnancy viability, estimate

gestational age, detect multiple pregnancies, and determine placental position [266]. While there is no compelling evidence that ultrasound scans reduce perinatal mortality [183, 69], they can be used to validate suspicious diagnoses without invasive and risky interventions, reduce labor induction for post-term pregnancy, and detect fetal malformation [266]. Moreover, a review of 58 obstetric articles, concluded that ultrasound imaging provides appropriate clinical management in at least 30% of cases when used by skilled operators [99].

Ultrasound imaging has also shown potential in the assessment of IUGR. A comparison between 38 IUGR and 32 appropriate for gestational age (AGA) fetuses showed that growth-restricted fetuses had a statistically significant thicker aortic wall than the AGA fetuses (1.9 mm vs. 1.15 mm) [57]. The median diameter of the abdominal aorta was also significantly higher in IUGR than in AGA fetuses. The thicker aortic wall in the IUGR fetus was also noted by [94], who compared 35 IUGR fetuses with 49 AGA fetuses. In contrast to [57], they reported on the substantially lower diameter of the abdominal aorta for fetuses with IUGR [94].

There is no scientific evidence for, or consensus on how often ultrasound scans should be performed during pregnancy. Some obstetricians recommend at least four ultrasound scans during normal pregnancies, whereas others recommend only one, to be performed before the 24th gestational week [191]. When four scans are performed during pregnancy, the first is conducted between weeks 10 and 14, to validate the pregnancy and estimate gestational age. The second scan is carried out between weeks 18 and 22, to detect fetal anomalies and confirm gestation age. The third scan is scheduled between weeks 30 34 of gestation, to assess fetal growth. The final scan is scheduled between weeks 36 and 38, and focuses on the fetal weight, position, and orientation/presentation, which helps to determine the optimal mode of delivery.

2.5.1 Fetal biometry

Ultrasound imaging enables measurement of different fetal organs, and estimation of gestational age and fetal weight. The most common measures are biparietal diameter (BDP), femur length (FL), head circumference (HC), crown rump length (CRL), and abdomen circumference (AC) [160]. Using a combination of these measurements, fetal weight estimates are within 5% of the actual weight in 50% of cases, and within 10% in 80% of the evaluations [83].

Fetal biometry measurements have been shown to be more accurate during the first trimester. During the second and third trimesters, fetal measurement accuracy is impacted by genetic and nutritional factors [202]. Recently proposed formulas combining the transcerebellar diameter (TCD) with FL and AC are emerging as a solution for dating late pregnancies (after the 24th gestation week), with gestational age estimates within $\pm$ 3 weeks of the CRL measurement taken between the 8th and 14th gestational weeks [64].

Fetal biometry measurements can differentiate between fetuses that are IUGR and those that are constitutionally small (SGA) [227]. Specifically, when an estimated weight is below the 10th percentile for gestational age, the fetus is considered growth-restricted, as defined by the American College of Obstetricians and Gynecologists (ACOG) guidelines [247]. However, a previous study, conducted by the Prospective Observational Trial to Optimize Pediatric Health (PORTO), found that only 2% of fetuses whose estimated birth weight was within the 3rd and 10th percentile, had an adverse perinatal outcome; the authors concluded that the threshold should be below the 3rd percentile [246].

Furthermore, fetal biometry measurements ignore the fact that 10% of the normal population is genetically predisposed to be small, thus increasing the false-positive rate [83]. Hence, to increase the accuracy in detecting IUGR, fetal biometry should be combined with methods assessing the fetal ANS physiology [83]. When IUGR is

detected, the pregnancy is categorized as high-risk, as the condition has long-term consequences.

2.5.2 Doppler velocimetry

Doppler velocimetry assesses the blood flow in the umbilical arteries and vein to evaluate pregnancies at risk of fetal compromise [242], such as growth restriction [6] or cardiovascular abnormalities [35].In healthy pregnancies, the placental and fetal circulation transfers oxygen and nutrients, and eliminates fetal waste products [160].

Umbilical flow is assessed using different indexes, such as systolic and diastolic ratio, pulsatility index and resistance index [13]. Higher indexes indicate significant vascular resistance, thus implying that fetal health is at risk [35, 221].

The resistance indexes are mainly measured on the umbilical artery (UA), the middle cerebral artery (MCA), and the ductus venosus (DV) [178]. Of these three areas, the UA Doppler is the only device that has been the subject of randomized controlled trials, which have supported its feasibility for fetal surveillance in high-risk pregnancies [7]. The UA Doppler measures the resistance in fetoplacental circulation flow, providing a pulsatility index (PI). In a healthy fetus, the UA has a forward flow. However, increases in placental resistance obliterate the muscular arteries in the placental villi, resulting in a reduced diastolic flow [35], which then eliminates and later reverses the fetoplacental circulation flow. Both the absence and the reversal of flows can be visualized in the Doppler images. In the case of the absent end-diastolic flow (AEDF), the pronounced systolic peak is followed by an interruption, while in the reversed end-diastolic flow (REDF), the systolic peak is followed by a negative peak. In fetal growth-restricted pregnancies with AEDF or REDF, delivery is recommended at week 32 [206].

Randomized controlled trials have demonstrated that a UA PI greater than the 95th percentile in restricted-growth fetuses is an indicator of a perinatal adverse

outcome [246, 190]. The use of UA Doppler was also shown to be effective in reducing the incidence of perinatal deaths and induced deliveries [7]. MCA flow can be used to detect problems caused by fetal hypoxemia in IUGR. In a hypoxic state, most of the oxygenated blood is supplied to the brain, heart, and adrenal glands, affecting the peripheral circulation [247]. This phenomenon is called *brain-sparing reflex* and is observable in the waveform of the MCA Doppler. MCA Doppler is also a reliable indicator of anemia. Moreover, the MCA PI/UA ratio can indicate adverse perinatal outcomes (Mari and Hanif, 2008), which are related to an increment of the diastolic flow due to hypoxia [180].

DV flow can be used to detect the cardiac failure in IUGR, particularly in cases of early-onset fetal growth restriction [27]. It is a reliable marker of acidemia and stillbirth [27], which are caused by absent or reversed end-diastolic pressure at the ductus venosus. Although DV flow measurement displays moderate accuracy in detecting fetal compromise, previous works have suggested that DV Doppler alone is insufficient for fetal surveillance [206]. Furthermore, DV Doppler does not offer any added benefit over traditional CTG for fetal monitoring [152]. Nevertheless, delaying delivery until finding an abnormality using DV flow could prevent neurological impairment in the long-term [85]. Randomized controlled trials are still needed to more accurately assess the benefits of DV flow measurement.

Other anatomical areas useful in the management of fetal growth-restricted pregnancies are the uterine artery, the aortic isthmus, umbilical vein, and the atrioventricular valves [178]. The uterine artery flow is useful in identifying pre-eclampsia and SGA neonates in high-risk pregnancies [206]. The aortic isthmus measures the balance between the brain's impedance and systemic circulation, indicating cardiac dysfunction when there is an abnormal balance [58]. Umbilical vein flow provides an indication of fetal venous circulation, where high values suggest increased venous pressure that results in right-sided heart failure and myocardial hypoxia [184].

2.5.3 Fetal echocardiography

Fetal echocardiography is a non-invasive ultrasonography technique that examines fetal cardiac anatomy and function [89]. The accuracy and speed of performing fetal cardiac assessment have improved in the last decades, following the introduction of advanced techniques such as color Doppler [67]. The primary use of fetal echocardiography is in the detection of congenital heart diseases (CHDs), which are the most common abnormality in fetuses, with a prevalence of around 8 to 9 per 1,000 live births [104]. The procedure assesses the heart structure, as well as the direction, pattern, volume, and velocity of flow [8]. The basic visualization of the chambers can be extended to include blood flow through the chambers, using a technique called 'five chambers views', which increases the sensitivity of detecting CHDs by 5%, achieving a final sensitivity rate of 65% [9].

Fetal echocardiography also includes a pulse wave Doppler component, which is recommended for a complete evaluation of the fetal heart. The pulse wave shows the blood flow through the atrioventricular, mitral, and tricuspid valves [1]. These valves generate a dual-peak Doppler waveform that comprises the E-wave, which is the passive diastolic filling, and the A-wave, which is the active diastolic filling ("atrial kick") [1]. In healthy fetuses, the amplitude of the A-wave is greater than that of the E-wave, which increases throughout gestation. A higher increase in the E-wave/A-wave ratio is a sign of IUGR or congenital cystic adenomatoid malformation, which can lead to mitral or tricuspid regurgitation [162, 159].

Modern echocardiography techniques include three-dimensional (3D) and four-dimensional (4D) fetal heart assessment [65], which enable real-time examination of the heart rate function, and a more accurate assessment of the heart structures [49, 33, 115].

Although fetal echocardiography is considered one of the most relevant fetal cardiac assessment techniques, it is costly and requires qualified specialists to perform the

examination [46]. Therefore, fetal echocardiography is only provided when indicated by specific maternal and fetal conditions.

2.6 Comparison of fetal cardiac monitoring methods

Table 2.1 presents an comparison of the fetal cardiac monitoring techniques presented in Sections 2.4 and 2.5, particularly with respect to the following four criteria:

1. medical equipment cost.
2. operator training requirement.
3. gestational week at which the device can be used.
4. evidence supporting the device's utility.

2.6.1 Cost analysis and availability in LMICs

The fMCG is an expensive method, requiring specialized operator training, dedicated shielded rooms, each costing approximately $350,000 to construct [232] and high maintenance [196, 44, 133], which have limited its use in HICs [44, 259] and its introduction into and widespread integration in LMICs.

Even compact portable ultrasound equipment, such as the GE LOGIQ Book XP (General Electrics, Milwaukee, WI, USA), costs at least $ 10,000, and carries additional expenses such as maintenance, supplies, battery replacement, and staff training [271]. However, a recent review on the use of ultrasound in LMICs reported an expanding utilization of low-cost, portable imaging technology in low-resource settings [230].

Table 2.1: Comparison of fetal cardiac monitoring methods. The first column presents a four-point ordinal scale of medical equipment cost, from low ($) to extremely high ($$$$). The horizontal line indicates when, during pregnancy, the technology can be used. The color of the line indicates the time required for training operators (green: low; blue: moderate; cyan: considerable; red: high; magenta: extreme). The thickness of the line indicates the relative evidence for the utility of each technology.

Cost	Mode	Stage in pregnancy: Antepartum, Gestational Week: 1-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	Intrapartum: Delivery
$	fPCG					*				
$	1D-Doppler ultrasound					†				
$	Hand-held Doppler					‡				
$$	CTG					§				
$$	Abdominal fECG				¶					
$$	Scalp fECG									‖
$$$	Ultrasound imaging					**				
$$$$	fMCG					††				

* GA $\geq$ 24 weeks [253].
† GA $\geq$ 20 weeks [196].
‡ GA $\geq$ 20 weeks [196].
§ GA $\geq$ 20 weeks [98].
¶ GA $\geq$ 18 weeks [207].
‖ Intrapartum (GA $\geq$ 36 weeks) [187].
** GA $\geq$ 20 weeks [273].
†† GA $\geq$ 20 weeks [196].

The ultrasound devices most commonly used for fetal monitoring were provided by Sonosite Inc (Bothell, WA, USA) [135, 97, 41, 140] and General Electric [188, 92]. Despite these early indications of their increasing integration in fetal monitoring in LMICs, there is still a need to assess the benefits, trade-offs, and potential drawbacks of large-scale obstetric ultrasound implementation in these regions [134].

Several randomized controlled studies using ultrasound imaging in antenatal care in LMICs, did not show any significant reductions in adverse perinatal outcomes [183, 69, 134, 92, 80]. Moreover, this technique requires specialized skilled operators,

who are in limited supply in LMICs [271].

Scalp fECG is limited to intrapartum use and requires specialized training. The most common device used for invasive fECG is the STAN monitor (Neoventa Medical, Goteborg, Sweden) [54]. [256] reported that the average cost of ST analysis in 2,827 deliveries in a Dutch hospital was € 1,345, which is not feasible for LMICs.

Abdominal fECG can be captured using low-cost equipment, which does not require skilled users [31, 174]. The Monica AN24 monitor (Monica Healthcare, Nottingham, UK) and the Meridian M100/M1000 monitors (MindChild Medical, North Andover, MA), both which have been approved by the Food and Drug Administration (FDA) and the European Commission (CE), are two commercial devices commonly used for non-invasive fECG [31]. However, although the cost of fECG devices is relatively low ($ 4,500 for the Monica Novii Wireless Patch System [45]), non-invasive fECG is still not widely used since the systems still require further testing to definitively demonstrate that the morphological analysis is similar to that provided by the scalp electrocardiography method [225, 123].

On average, CTG machines cost at least $ 450 [274] and require maintenance, supplies, and training, thereby limiting its use in low-resource settings [273]. Although obstetric protocols in HICs recommend CTG, its use in LMICs has not improved fetal outcomes in comparison to auscultation methods [108]. In contrast, auscultation methods, particularly the fetoscope, have been shown to be associated with reduced perinatal deaths in LMICs [260].

Of the various auscultation methods available, the Pinard stethoscope is the most available tool in resource-constrained regions due to its affordable cost [118]. Before its introduction, midwives used a stethoscope in the labor ward to listen to the fetal heart rate for ten minutes every half hour [158]. However, auscultation with the stethoscope can be uncomfortable for both patients and practitioners. Additionally, the stethoscope provides unsatisfactory results due to confounding factors such as

environmental noise [158].

Hand-held Doppler transducers are simple to use and can be purchased for as little as $17 [237]. They require less time than the Pinard stethoscope to assess the fetal heart rate [158]. [198] compared the performance of hand-held Doppler and Pinard devices for fetal monitoring in the intrapartum period by reviewing 19 studies conducted in India and African LMICs. The comparison showed that Doppler devices accurately detected more fetal abnormalities than the Pinard stethoscope. However, there was no statistically significant improvement in perinatal outcomes when an anomaly was detected. The authors suggested that the lack of progress resulted from the poor clinical management and protocol referral of abnormality events. The review also found that both patients and medical providers preferred Doppler hand-held devices than Pinard stethoscopes, thereby justifying the integration of Doppler in fetal monitoring protocols for LMICs.

2.7 Usage of devices in LMICs

In LMICs, there is limited availability of the life-saving and complex medical devices routinely used for fetal cardiac monitoring in high income countries. The technologies involve the use of ultrasound technology, telemedicine, CTG and other fetal monitoring techniques that may not be easily implementable in low-resource settings. However, multiple research studies have demonstrated the acceptable effectiveness of suitable and appropriate technologies in these regions for fetal cardiac monitoring [26, 280, 200, 66, 11].

Mdoe *et al.* [175] demonstrated the superiority of fetal heart rate monitoring using continuous Doppler when compared to intermittent fetoscope auscultation. In their study, there was an 8.1% incidence of abnormal fetal heart rate detection when using continuous Doppler, versus 3.0% with the use of intermittent auscultation.

Additional studies have provided evidence for the preferential use of Doppler ultrasound for fetal cardiac activity monitoring. It has also been shown to accurately determine the number of reported fetal demise and to classify them as stillbirths versus neonatal deaths [155]. Additionally, the emotional reassurance among mothers associated with hearing the fetal heartbeat amplified by the Doppler is a positive experience that could influence positive outcomes. Compared to the Pinard stethoscope, the Doppler was noted to be superior in detecting abnormal intrapartum fetal heart rate but was not associated with improved perinatal outcomes [275].

Britton *et al.* [42] demonstrated the use of 'tele-ultrasound' for fetal cardiac monitoring in low resource settings. In the study, fetal monitoring devices were delivered to participants, to overcome challenges of geographical distance, lack of facilities and inadequate healthcare personnel that are common in these areas. The 'tele-ultrasound' technique was found to be low-cost, and reliable and implementable in resource-limited settings.

[276] reported on improved perinatal outcomes in following the use of CTG during labor in LMICs. Others have associated CTG (cardiotocography) with higher Cesarean section rates, with no added benefit on perinatal outcomes, and therefore do not recommend its use in low-resource settings [109]. High-quality evidence considering implementation barriers and enablers is needed to determine the optimal fetal monitoring technique in Low-resource settings. [276] noted that there are significant gaps between international recommendations and what is practically possible in most resource-constrained countries.

A study in Uganda highlighted critical challenges as shortage of staff and devices, institutional challenges, and maternal perceptions to monitoring [181]. Another study in Tanzania listed lack of strict protocol for use and misidentification of maternal heart rate as the challenges associated with the introduction of Moyo, an electronic strap on a fetal heart rate monitor [141] developed to improve intrapartum fetal heart rate

monitoring. More specifically, the quick, short, and unstructured assessments and inferences limited the initiation of various interventions due to indecisiveness. The introduction of CTG was associated with simpler and more efficient monitoring of labor, but no improved outcomes [125, 141].

2.8 Telemonitoring for perinatal care, an alternative for LMICs

In recent years, telemonitoring applications have been developed to enhance maternal and fetal monitoring. These applications have been made possible by the high penetration of mobile telecommunications technologies in LMICs, with approximately 90% of the population owning a mobile phone device [34]. This high coverage of mobile phone access can be exploited to overcome perinatal care access barriers in LMICs, such as low literacy levels, poor road infrastructure, and medical professionals and equipment deficiencies [73, 226, 102].

The feasibility of mobile health applications (mHealth) in improving antenatal care was presented by [75] after reviewing 14 cases conducted in sub-Saharan Africa, Southeast Asia, and Middle-East countries. The authors found that mHealth solutions can improve perinatal care services by increasing the percentage of women attending the minimum WHO-recommended perinatal visits. They noted that the most effective mobile apps were those that used client education and behavior change communication via short messages and patient tracking to allow for patient follow-up in subsequent visits.

In a review of telemonitoring in obstetrics, researchers reported that mobile applications connected to external devices, such as electrodes, body sensors, and thermometers provided effective maternal and fetal monitoring [11]. The used external devices enabled the digitalization of data, which was later analyzed by medical pro-

fessionals or artificial intelligence techniques to detect abnormalities. Among the available fetal monitoring mobile applications, [117] developed a method to identify the location of the fetus from an ultrasound image. The algorithms tested with pictures taken in a public hospital in Indonesia, demonstrated 93% accuracy in the detection of the fetal head and abdomen. Similarly, [132] developed methods to calculate the mean abdominal diameter (MAD) from an ultrasound image. A mobile app prototype was tested on ultrasound images captured by professional midwives in a Norwegian hospital, and demonstrated a mean error of -0.06 mm.

Awiti *et al.* [21] developed an Android-based digital fetoscope prototype. Fetal heart sounds are acquired using a Pinard horn and a microphone and are later sent to a smartphone via Bluetooth technology. In the smartphone, the audio signals are processed to display a heartbeat. When testing the system on adults, and comparing the measurements with those of a standard electronic sphygmomanometer, the Android-based digital fetoscope achieved a root mean square error of 7.23 bpm with a standard deviation of 5.44 bpm.

Tapia-Conyer *et al.* [239] introduced a mobile maternal-fetal monitoring application in a resource-poor and educationally-limited community in Mexico. The project aimed to evaluate the feasibility of providing remote antenatal care. The staff in the rural medical center was trained to use the mobile fetal monitor, which comprised a fetal ultrasound heart monitor, a uterine tocodynamometer, and additional tools for recording maternal blood pressure, blood glucose, and urinary protein values. The researchers split the 125 volunteers into control and study groups. The study group received perinatal care at the local medical center using the mHealth system, whereas the control group received standard perinatal care at the main public hospital. [239] observed that volunteers using the mHealth system were more than twice as likely to adhere to antenatal care monitoring than those receiving standard of care. There were no statistically significant differences in adverse perinatal outcomes between the two

groups, suggesting that the tested mobile technology did not compromise maternal and fetal health.

A low-cost fetal monitoring system introduced in a rural Guatemalan community in 2013, was later developed into a system that was tested in a randomized controlled trial in 2015-2017 [237, 166, 165]. The mHealth system consists of a low-cost one-dimensional Doppler transducer and a blood pressure device connected to an Android-based smartphone running an app designed for low-literacy traditional birth attendants (TBAs). The TBAs were trained to use the mHealth system during home visits. When a TBA visited a patient, the app guided the TBA to find the fetus and record a Doppler recording of the fetal cardiac activity for up to 20 minutes. The TBA also recorded blood pressure readings from both arms. The app then guided the TBA through basic questions presented through appropriate pictograms and audio prompts, to assist in the identification of concerning signs and symptoms during pregnancy. In the event a risk factor was identified, the app connected the TBA to appropriate (local or remote) medical care through a voice call, to provide decisional support and onward referral to appropriate healthcare, if needed. The Doppler signal and maternal blood pressure recorded with the mHealth system has allowed the development of different modules for providing estimates of fHRV, gestational age, and hypertension [250, 251, 252, 248, 249, 130].

2.9 Discussion and conclusion

Fetal monitoring is performed with a variety of devices and approaches, with CTG and ultrasound imaging considered the 'standard of care' in high-income countries. Despite the paucity of evidence supporting the utility of these techniques in reducing perinatal mortality and morbidity [183, 66], their use may still be beneficial throughout pregnancy. Specifically, CTG may facilitate the detection of signs of hypoxia re-

quiring Cesarean delivery [98], and ultrasound imaging can help estimate gestation age before gestational week 24, and to detect multiple pregnancies and fetal abnormalities requiring vigilance or/and interventions [99, 266]. However, in resource-constrained settings, CTG and Doppler imaging are scarce, due to their high cost and the need for trained operators [69].

Among fetal monitoring techniques, 1D Doppler transducers provide an affordable option, with an outstanding balance between cost, clinical utility, and operator learning curve, thereby making its use most practical (see Table 2.1). Moreover, Doppler transducers have been shown to be comparable to CTG in fetal heart rate monitoring in LMICs. Taken together, Doppler transducers, which are widely available in these regions [28], are a reliable 'standard of care' in low-resource regions [66, 108].

Perinatal care access in LMICs can be facilitated by telemetry [34]. Mobile technology can support fetal monitoring analysis and transmission of clinical information collected using low-cost devices, such as Doppler transducers, portable CTG, or auscultation methods. In this manner, economic and geographical barriers can be overcome to increase perinatal care coverage in LMICs.

The feasibility of the use and impact of mHealth mobile applications in fetal monitoring has been shown in several works conducted in LMICs [239, 238, 75, 165]. With the advent of increasingly complex smartphones, particularly those with embedded 'AI' chipsets), mHealth applications may extend beyond data collection and decision support systems, by processing complex maternal and fetal information in an edge computing paradigm, which allows the use of mobile applications without relying on network communication. However, regulation (especially in the US) is likely to limit this development, and to drive the solutions to higher-cost, self-contained devices.

Increasingly higher bandwidth cellular networks in LMICs could mitigate this by driving the processing to the cloud. Still, the economics of providing high-bandwidth networks to the poorest and least populated parts of the globe will still leave signif-

icant disparities in fetal monitoring care. A store-and-forward approach may offer a reasonable solution to this, with text messaging and voice calls addressing immediate issues and use of a robust remote decision-support network of trained professionals and an integrated referral mechanism (see [237] for example)

In summary, mHealth systems bear a significant potential to provide remote perinatal care and prevent many fetal complications, by removing many barriers prevailing in LMICs. Notably, such approaches can empower the frontline healthcare workers (and perhaps even mothers) to learn, and even improve the systems, and ultimately alleviate the "brain drain" in the medical field in LMICs [56].

Chapter 3

Data collection and labelling [2]

3.1 Abstract

The scalability of medical technology in low resource settings requires a higher level of usability and clear decision support compared to conventional devices, since users often have very limited training. In particular, it is important to provide users with real time feedback on data quality during the patient information acquisition and improve the design by including machine learning algorithms to automate fetal maternal health assessment. In this work, We present a preprocessing and labeling approach to prepare the data for further analysis which ultimately will run on-device and interact with the healthcare workers or mothers directly.

3.2 Introduction

This work has focused on reducing the high burden of perinatal deaths using a smartphone-mediated affordable perinatal screening system, which addresses many cultural requirements for use in rural Guatemala [237, 166]. The system allows the

[2]

monitoring of fetal heart rhythm, which is commonly used by clinicians to identify non-reassuring fetal status for timely intervention [22]. A low-cost One-Dimensional Doppler Ultrasound (1D-DUS) transducer, connected to a smart phone, was introduced into rural communities in collaboration with Wuqu' Kawoq — Maya Health Alliance, an NGO aiming to provide health care solutions for Guatemala's indigenous communities. Indigenous traditional birth attendants (TBAs) were trained to use the approach for monitoring fetal wellbeing during pregnancy [233, 237, 166].

The mHealth technology has provided TBAs with decision support, and through cellular connectivity, has linked their traditional procedures with a formal health-care referral process. Although the technology has so far proven effective [235, 233, 237, 166], the need of some refinements has been identified. One critical requirement is to ensure the quality of the 1D-DUS recordings. Indeed, during the first two release cycles of the app, around 40% of the recordings were low quality [166]. Despite the quality improvements by retraining the birth attendants and fixing the device connections, the signal quality had to be automatically evaluated to identify users who are making habitual mistakes or to identify equipment malfunctions. The low quality of recordings can distort the posterior fetal health analysis and complicate fetal abnormality detection. Since fetal heart rate (FHR) analysis is key to detection of IUGR in our population, and its accuracy depends on the DUS quality, the exclusion of poor quality DUS records before performing any analysis is important for reducing false positives [236, 156].

To evaluate the performance of automatic FHR estimation algorithms, it was necessary to manually annotate the heart rate in each database. This was performed on a temporal sequence of overlapping 3.75 s windows of 1D-DUS data. Three datasets were used in this work including Guatemala RCT Database, Leipzig University Hospital database and Oxford JR Database.

3.3 Methods

3.3.1 Databases

3.3.1.1 Guatemala RCT database

This dataset was collected as part of a randomized controll trial, conducted in rural highland Guatemala in the vicinity of Tecpán, Chimaltenango. The study focused on the use of the Doppler device, and an accompanying app with data capture and decision support software built-in, to improve the continuum of care for indigenous women of the target region. The study was approved by the Institutional Review Boards of Emory University, the Wuqu' Kawoq I Maya Health Alliance, and Agnes Scott College (Ref: IRB00076231 - 'Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala') and registered as a clinical trial (ClinicalTrials.gov identifier NCT02348840). All 1D-DUS signals were recorded by traditional birth attendants (TBAs), who were trained to use the hand-held device. Before recording the signals, the TBA also entered the gestational age in months and the maternal heart rate, measured using a self inflating blood pressure device (Omron M7), into the same mobile application designed to record the 1D-DUS.

3.3.1.2 Leipzig university hospital database

This dataset, used for training the FHR estimation algorithm in this study, was collected at the Leipzig University Hospital (LUH) in Germany, as part of the study presented in [17]. The database included data from 16 volunteers with pregnancies between the 20th and 27th week of gestation, including pathological cases such as Inter-uterine growth restriction (IUGR), premature rupture of membranes, or fetal heart failures. The study was approved by the Leipzig University Hospital ethics committee (record 348-12-24092012), and written informed consent was obtained from each patient. For each subject, indirect abdominal fetal electrocardiogram (fECG), a

maternal ECG reference and a 1D-DUS signal were simultaneously recorded by clinicians. The fECG recordings were acquired from 7 abdominal channels using a 16bit commercial ADC using the ADInstruments ML138 Octal Bio Amp and ADInstruments PowerLab 16/30 (ADInstruments, Dunedin, NZ), and stored at a sampling frequency of 1000 Hz. Spectral filtering was also performed in the hardware by a mains filter (cutoff frequency at 50 Hz) and a first-order high-pass filter (cutoff at 1 Hz).

3.3.1.3 Oxford JR database

This dataset was collected at the John Radcliffe (JR) Hospital in Oxford, UK. The study was approved by the NHS Health Research Authority (REC reference: 12/SC/0147) and written consent was obtained from each study subject prior to data collection. Each subject received detailed information on the study protocol and their right to withdraw from the study at any stage of the recording session, which was carried out by professional midwives. The dataset included 1D-DUS signals from 17 healthy pregnant women, who bore singletons between 20 and 38 weeks of gestation. This database has also been used in previous related 1D-DUS studies [236, 234, 250].

3.3.2 Data labeling and annotation

3.3.2.1 Annotation of Guatemala RCT database

Three independent annotators listened, visually inspected and labeled 195 DUS recordings from Guatemala RCT Database using a Matlab (MathWorks, Natick, MA, USA) graphical user-interface (GUI) interface (Figure 3.1). The GUI split recordings into segments of 0.75 s (seconds), allowing to annotate each of them as one of the following six categories:

- Interference: The epoch contains electrical interference, typically manifesting

as sort bursts of a buzzing sound.

- Silent: The epoch is silent or is barely audible.
- Talking: The epoch may contain audible heart beats but also human voices or noises from the environment generated by animals (e.g. dog barking).
- Poor quality: The epoch contains noise but not any of the other classes.
- Unsure: The epoch contains a mixture of sounds, which was challenging to assign a specific class, or the annotator was unsure of to which class it belonged.
- Good quality: The epoch contains audible heart beats with no significant presence of any of the above categories.

In previous works, the window for assessing DUS quality had been fixed at 3.75 s because it is the usual length for computerized analysis of fetal non-stress tests based on the Dawes/Redman criteria [192, 62]. However, in this work, the DUS quality was assessed for different window lengths to find which is more appropriate for this aim. Thus, after quality annotation, five different possible segments were built using as window length a multiple of 0.75 s; namely, the defined window length were: 0.75, 1.50, 2.25, 3.00, and 3.75 s. In order to identify such windows, only contiguous windows of a given class were used to create the windows longer than 0.75 s, creating fewer examples for longer windows. To maximize data availability, one 0.75 s could appear more than once in a longer window.

After labeling the quality of the signals, segments that had been manually classified as good quality were used for beat annotation. Since in Doppler ultrasound each cardiac cycle is represented by a combination of cardiac wall and valve movements [219], it is extremely complicated to mark one specific point as a beat location, thereby producing a large variation among annotators. Listening to the data and attempting to hit a button when a beat is heard is also problematic since human

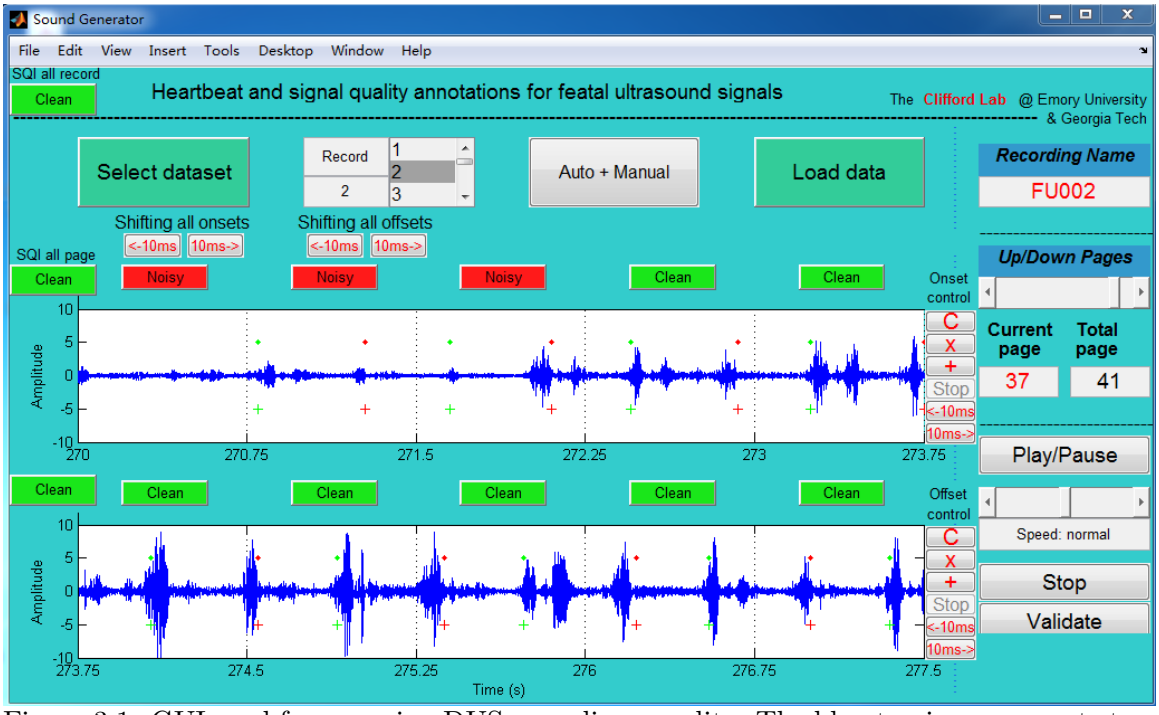


Figure 3.1: GUI used for assessing DUS recordings quality. The blue tracing represents two contiguous strips of 3.75 s raw audio file, each broken into five 0.75 s segments. The entire 7.5 s segment could be played back in real time or at fractional speeds (with pitch-preserving frequency shifting), paused or looped. The green and red crosses indicate the start and end of each 'heart beat' as determined by an automated algorithm [233], which were used only for guidance. Using 'Sennheiser HD 202 II Professional' headphones, each segment of 0.75 s was labeled by three trained researchers, acting independently, as either good quality, poor quality, interference, silent, talking, or unsure.

reactions, keyboard delays, etc., add in large variable time delays [237]. To address this problem, we designed a Matlab GUI (Figure 3.2) to count the number of audible beats in each 3.75 s segment. The beat counting was performed by three independent annotators. The median number of beats over all three annotators, b, was used to define the FHR estimate as $FHR = 60b/3.75$ BPM.

3.3.3 Annotation of the Leipzig university hospital database

For the Leipzig university hospital database, the fECG channels were visually inspected to locate beats in both the Doppler and fECG recordings. Since the fECG was recorded from the maternal abdomen, the first step was to remove the maternal components. To do this, a previously validated fECG extraction method based on an

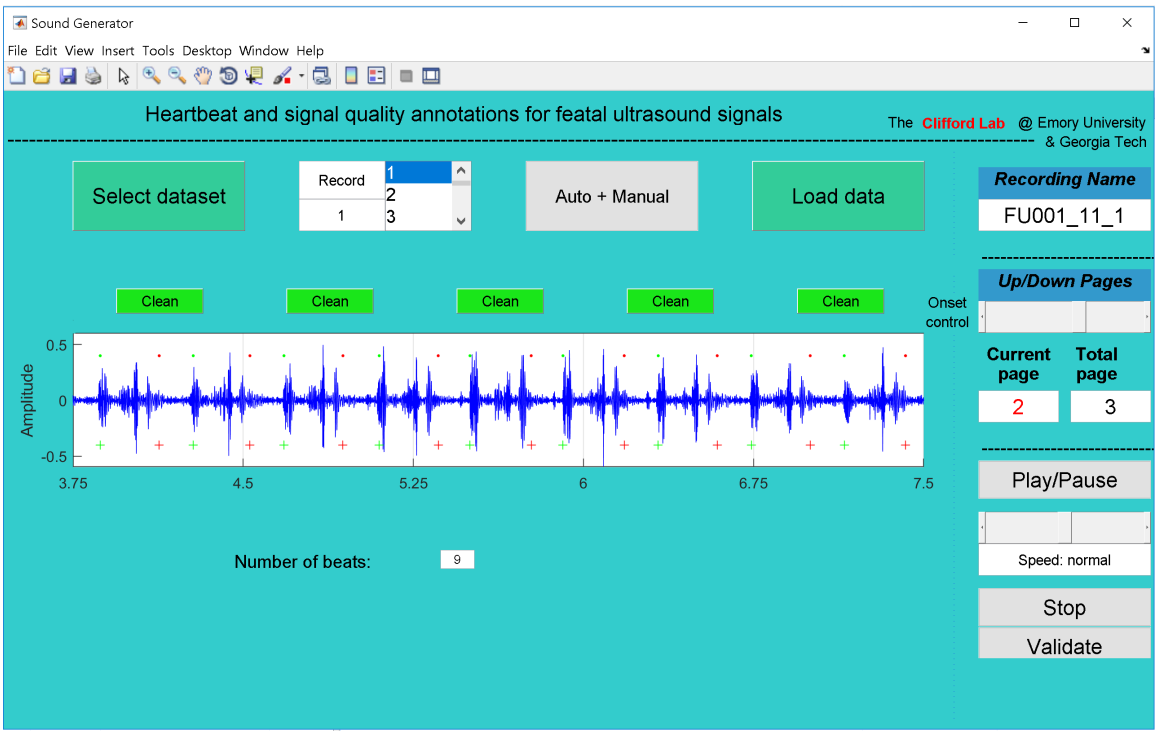


Figure 3.2: GUI used to manually annotate the number of beats in the Guatemala RCT dataset. Annotators listened to each 3.75 s segment, counting and recording the number of audible beats. © Institute of Physics and Engineering in Medicine. Reproduced with permission. All rights reserved (`https://iopscience.iop.org/article/10.1088/1361-6579/ab033d`).

extended Kalman smoother was used [30, 210]. Then, the filtered fECG and the 1D-DUS signal were resampled to 4 Khz, and were displayed in a graphical user-interface (GUI) (Figure 3.3), using a window size of 3.75 s, written in Matlab (MathWorks, Natick, MA, USA). The 3.75 s window was chosen because it is the usual length for computerized analysis of fetal non-stress tests based on the Dawes/Redman criteria [192, 62]. Furthermore, this window length was shown to be suitable for assessing 1D-DUS quality acquired with the same hand-held device used in this study [251]. In addition to the fECG and 1D-DUS signals, the Matlab GUI also displayed the estimated times of the QRS peaks from both maternal and fetal ECG using algorithms in the FECGSYN toolbox [30]. These estimated fECG QRS peak times were taken as guide for locating the beats in the 1D-DUS signal.

Two independent annotators used the Matlab GUI to assess the quality of 1D-DUS and fECG channels, and to place the beat time location based on fECG channels.

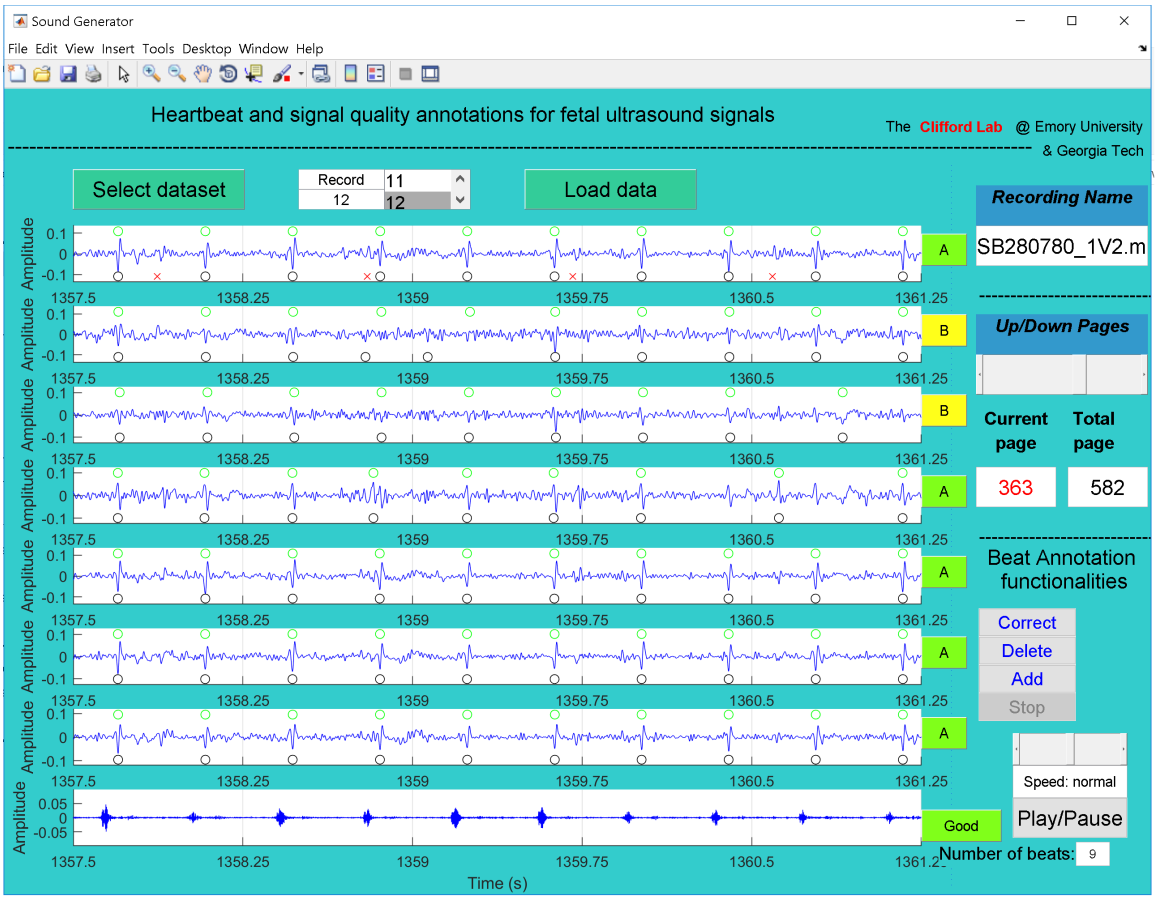


Figure 3.3: GUI used to annotate LUH database. The first seven channels correspond to abdominal fECG after filtering maternal components using an extended Kalman smoother and peak detection of fetal (black circles) and maternal beats (red crosses, upper plot). The last channel in the plot is the 1D-DUS signal re-sampled at 4 kHz. Using buttons on the right of the GUI, and the automatic detection as a preliminary guide, two independent annotators provided quality of fECG and 1D-DUS. For good quality 1D-DUS, fECG beats are located (green circles in the upper part of each fECG subplot).

For each 3.75 s segment, annotators listened to the ultrasound recording and noted the number of audible beats, and labeled the 1D-DUS quality using the same class hierarchy described in [251], namely, good, poor, electrical interference, talking, silent, or unsure. Since 1D-DUS quality may affect the FHR estimation [234], only 1D-DUS segments with good quality were retained for heart rate estimation. After labeling the 1D-DUS quality, annotators labeled each fECG channel as:

- A: All QRS complexes can be seen (although not necessarily in the same channel)

- B: Some QRS might be missing or some extra beats
- C: Lots of noise and absent signal/dropout but see at least two neighboring beats
- D: Almost completely noise
- E: Unsure

To annotate the beat time location, the visible peaks contained in the good quality fECG channels were used. As an initial estimate, the location provided by automatic fECG QRS detection was used; however, annotators were able to correct those locations using the GUI. To avoid confusing maternal breakthrough for fetal peaks, annotators used visual inspection of the maternal ECG and detected peaks, provided in the upper subplot of the GUI (red crosses in Figure 3.3), thus discarding any peak when it was aligned to a maternal peak and out of sequence. Observation across all fECG channels was used to improve the accuracy of beat time locations.

After finishing the annotation process for all the 1D-DUS and fECG channels and retaining segments with simultaneous high quality fECG and 1D-DUS, 5 of the 16 subjects were included, the remaining were eliminated due to high noise levels in either of the channels. (Data were collected serendipitously as part of another study in which 1D-DUS recording quality was not prioritized.)

To ensure that beat time locations were consistent, the difference in seconds, δ, of fECG peak times between the two annotators was compared. Figure 3.4 shows that for 95% of annotated beats, the difference between pairs of annotations was less than 50 ms. Therefore, a high level of trust was ensured in the fECG annotations.

3.3.4 Annotation of Oxford JR database

Each of the 1 minute-length 1D-DUS signals in this dataset were labeled by three different expert annotators using a Matlab GUI. Each reviewer independently labeled

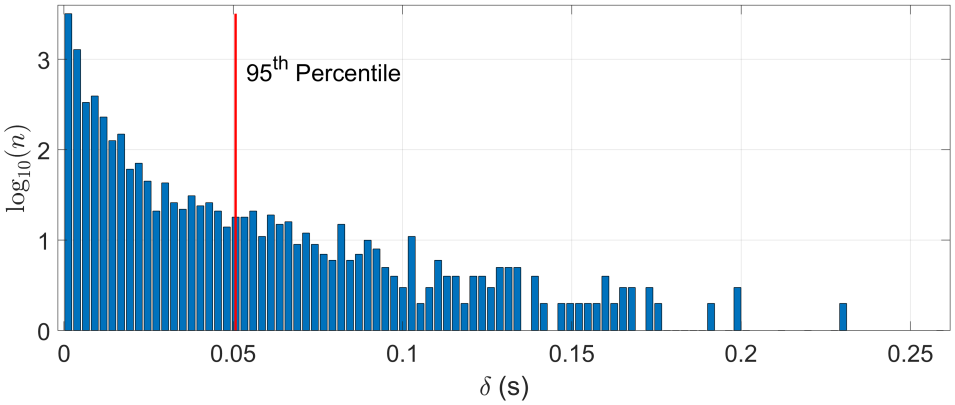


Figure 3.4: Histogram of time differences between beat time locations from two independent annotators. The horizontal axis, δ, is the difference in seconds of the two annotations for the fECG peak timing. The vertical axis represents a logarithmic scale of the count (n) of each difference. Note that 95% of the annotations differed by at most 0.05 s.

the quality of each second as:

- Noise: No information available in the signal.
- Poor: The signal may contain heartbeats, but it is too 'noisy' to identify them.
- Intermediate: Difficult to hear heartbeats, but can be done with some effort. Heart rhythm detection may be possible.
- Good: Some background noise, but heartbeats can be heard clearly. Heart rhythm detection is possible.
- Excellent: Almost no background noise, heartbeats are easy to identify, heart rhythm detection is possible.

These categories defined the signal quality of the 1D-DUS segments. Furthermore, while annotators were listening to the 1D-DUS segment, they clicked a mouse to indicate the temporal location of each beat that they heard.

After labeling all the segments, one-minute segments were split into 3.75 s with no overlap. Only segments in which at least two annotators labeled the same class were

used. The manual FHR was estimated by first aligning the points that annotators clicked for the beat sound. The closer points were grouped and their median was taken as initial beat location. These locations were corrected using the homomorphic envelope of the 1D-DUS segment. Starting from the last beat location of the segment, the closest peak was searched in a window starting one interval prior to the annotation and ending 1/4 interval thereafter. Then, using reverse iteration, each peak time was corrected by finding the maximum peak in a window of $\pm$ 15 BPM from the last corrected peak. More description of the method can be found in [234].

Similar to the other datasets, for the Oxford JR dataset, the manual FHR was estimated as $FHR = 60/median(I)$ BPM, where I is a vector containing the difference in seconds between two corrected adjacent peaks.

3.4 Discussion and conclusion

In this work, the description of the datasets and the labeling process are provided. The Guatemala RCT database was captured by TBAs in rural areas of Guatemala with minimal training in the use of the equipment (a low-cost Doppler device connected to a smartphone). This dataset is used in chapters 5 and 7 for gestational age estimation using deep learning approaches.

The Leipzig database was recorded at the Leipzig University hospital in Germany. In this database, in addition to 1D-DUS, simultaneously recorded fetal ECG signals were provided. Another database used in this study is the Oxford JR database collected at the John Radcliffe hospital in Oxford and included the 1D-DUS signals. These two databases are used to evaluate the segmentation model presented in chapter 4.

Chapter 4

Unsupervised fetal Doppler signals segmentation and heart rate variability estimation [3]

4.1 Abstract

One dimensional (1D) Doppler ultrasound (DUS) is commonly used for fetal health assessment, during both regular prenatal visits and labor. It is used in preference to ECG and other modalities because of its simplicity and cost. To date, all analysis of such data has been confined to a smoothed, windowed heart rate estimation derived from the 1D DUS signal, reducing the potential of short-term variability information. A first step in improving the assessment of short-term variability of the fetal heart rate (FHR) is through implementing an accurate beat detector for 1D DUS signals.

This work presents an unsupervised probabilistic segmentation method enabled by a hidden semi-Markov model (HSMM). The proposed method employs envelope and spectral features for an online segmentation of fetal 1D DUS signal. The beat onsets

and fetal cardiac beat-to-beat intervals are then estimated from the segmentations. For this work, two data sets were used, including 1D DUS recordings from five fetuses recorded in Germany, comprising 6521 beats and 45.06 minutes of data (dataset 1). Simultaneous fetal ECG (fECG) was used as the reference for beat timing. Dataset 2, comprising 4044 beats captured from 17 subjects in the UK was hand scored for beat location and was used as an independent held-out test set. Leave-one-out subject cross-validation was used for parameter tuning on dataset 1. No retraining was performed for dataset 2. To assess the performance of the beat onset detection, the root mean square error (RMSE), F1 score, sensitivity, positive predictivity (PPV) and the error in several standard common heart rate variability metrics were used. These metrics were evaluated on three fiducial points: 1) beat onset, 2) beat offset, and 3) middle of beat interval.

In dataset 1, the proposed method provided an RMSE of 20 ms, F1 score of 97.5 %, a Se of 97.6%, and a PPV of 97.3%. In dataset 2, the proposed method achieved an RMSE of 26 ms, an F1 score of 98.5 %, a Se of 98.0 % and a PPV of 98.9 %. It was also determined that the best beat-to-beat interval was derived from the onset of each beat. For the dataset 2, significant correlations were found in all short term heart rate variability metrics tested, both in the time and frequency domain. Only the proportion of successive normal-to-normal interval differences greater than 20 ms (pNN20) exhibited a significant absolute difference.

This work presents the first-ever description of an algorithm to identify cardiac beats with 1D DUS, closely matching the fetal ECG-derived beats, to enable short-term heart rate variability analysis. The novel algorithm proposed requires no human labeling of data, and could have applicability beyond 1D DUS to other similar highly variable time series.

4.2 Introduction

There are 2.9 million annual neonatal deaths (deaths in the first 28 days after birth) worldwide [245], and the biggest risk factor is the small size of the child at birth due to preterm birth or issues leading to fetuses being small-for-gestational-age (SGA) [147]. Growth retardation also increases the risk of post-neonatal mortality, growth stunting, and adult-onset non-communicable diseases. The burden is inequitably carried by low-and-middle-income countries where 99% of intrapartum stillbirths occur [150]. Although many resources have been leveraged to alleviate this problem, progress has been limited. The failure of intrapartum monitoring to improve outcomes is becoming increasingly understood, and improved fetal monitoring strategies are needed in order to reduce the incidence of intrapartum-related stillbirths and neonatal deaths [212].Therefore, it is essential to provide a low-cost, accessible and accurate health care system for this purpose, targeting patients who have difficulties in accessing appropriate health care services.

Fetal monitoring is commonly performed with cardiotocography (CTG) devices. CTG uses one dimensional Doppler ultrasound (1D DUS) to estimate fetal heart rate (FHR), which is further analyzed to identify fetal risk. However, regardless the widespread use of CTG, randomized controlled trials (RCTs) have reported little improvement in perinatal outcome, including intrapartum stillbirth or neonatal deaths [60, 87, 43, 126]. In fact, the use of CTG has not had a statistically significant impact on reducing perinatal death in comparison to the use of more traditional methods, such as Pinard stethoscope or fetoscope [127, 176]. Perhaps a key weakness of the use of Doppler for estimating FHR is that CTG estimates FHR using conventional autocorrelation techniques that average the heart rate over several seconds [252]. This leads to a loss of detailed information. Never-the-less, in recent works, we have demonstrated that the fetal heart rate variability (FHRV) from 1D DUS can be used to identify intrauterine growth restriction (IUGR) in antepartum monitoring [238].

We subsequently hypothesized that the analysis of FHRV derived from raw 1D DUS can be improved by more accurately identifying beat-to-beat intervals, and collecting multiple data sets in order to test this hypothesis [165]. Following this hypothesis, we propose an unsupervised method for identifying beat-to-beat timings based on hidden Markov models (HMMs) consisting of a clustering process for the recognition of states, followed by probabilistic segmentation to detect the onset of beats. Metrics for success include onset detection performance and the correlation of FHRV estimates with reference values.

4.3 Background

4.3.1 Clinical trials using FHRV

Randomized controlled trials (RCTs) of 1D DUS in high-risk pregnancies were first reported in the 1990's to study the effect of 1D DUS monitoring. In 1992 Davies *et al.* [60] conducted an RCT on a general obstetric population of 2475 pregnant women with singletons between 19 and 22 weeks gestational age. The intervention group underwent routine screening of the umbilical-artery and uterine-artery with DUS according to their risk of having an SGA fetus. The authors reported that this intervention did not demonstrate any improvement in neonatal outcome. In 2003, Giles *et al.* [87] reported that a close surveillance in twin pregnancy (526 women at 25 weeks of gestation) resulted in a lower than expected fetal mortality in both the no Doppler and Doppler groups. In 2014, an RCT was performed on 1971 pregnant women in Kampala comparing wind-up fetal Doppler versus the use of the more traditional (non-electronic) Pinard stethoscope for intermittent FHR monitoring in labor [43]. They concluded that routine monitoring with a hand-held Doppler could increase the identification of FHR abnormalities in labor. However, their trial did not find evidence that this leads to a decrease in the incidence of adverse events, including

intrapartum stillbirth or neonatal death. A similar follow-up study on 2844 women in 2018 in Dar es Salaam indicated that abnormal FHRs were more often detected in the Doppler (6%) versus Pinrad (3.9%) [127]. Once again, the overall findings did not demonstrate any improvement in perinatal outcome. The same authors reported results (on potentially the very same study) on 2442 pregnant women in Tanzania to identify abnormalities, defined as FHR<120 or FHR>160 bpm [126]. The secondary outcomes were rates of assessment/documentation of FHR, obstetric time intervals, and intrauterine resuscitation. The authors report that Caesarean section rates significantly increased from 2.6 to 5.4%, and vacuum deliveries significantly increased from 2.2 to 5.8% (both with $p < 0.001$). Perinatal outcomes i.e., fresh stillbirths and early neonatal deaths were not significantly different. The study was limited by both lack of randomization, matched populations, and involvement of low-risk pregnant women with fewer adverse perinatal outcomes than would be expected in a high-risk population. Another comparative FHR monitoring study, comparing intermittently used fetoscope and hand-held Doppler that took place in rural Tanzania failed to demonstrate a statistically significant difference in the detection of abnormal FHR between intermittently used Doppler and fetoscope and adverse perinatal outcomes [176].

4.3.2 Fetal heart rate variability estimation from Doppler

Over the last decade or more, there have been several attempts to improve FHRV estimates from 1D DUS signals. Studies on enhancing fast Fourier transform (FFT) technique are among the first approaches to incorporate beat-to-beat variability estimation from 1D DUS. Parametric spectral analysis using the autoregressive model[79] and the multiple signal characterization algorithm [74] had been used to overcome the shortcomings of FFT based approaches. Another category of studies had been presented on developing the correction algorithm to enhance autocorrelation based

methods. The correction methods include removing the constant error, which has been assigned to an averaging nature of the autocorrelation function [278], and dynamic adjustment of the autocorrelation window and peak detection algorithm [120]. The most recent study on assessing FHRV using 1D DUS processing has employed the empirical mode decomposition (EMD) and the kurtosis statistics [3]. The hybrid EMD-Kurtosis approach showed a better estimate of mean beat-to-beat in comparison to the autocorrelation method. However, none of the methods described so far provide a satisfactory evaluation of the FHRV estimates in terms of clinical utility.

4.3.3 Beat segmentation in the 1D DUS

The method for 1D DUS beat onsets detection proposed in this work is based on an analysis of sequential time series and probabilistic segmentation. The HMM and its extensions are one of the most popular methods in this field, and several approaches to the segmentation of cardiac signals using this technique have been published. Koski [138] was perhaps the first to describe the application of an HMM to segmentation of cardiac time series (specifically the electrocardiogram). Clavier, Graja, and Boucher extended this approach by using wavelet basis functions [52, 96]. Hughes et al. then extended this approach and compared HMM with hidden semi-Markov model (HSMM) [114, 113]. HMMs were also used for automated identification of fetal heart valve movements from 1D DUS recordings [173, 172], which requires simultaneous fECG as a reference. Recent work on phonocardiograms (or heart sounds) using HMMs have shown great promise for segmenting cardiac time series [215, 228]. Most recently, Stroux and Clifford applied a similar approach to 1D DUS [234]. Although this approach showed promise, the high variance of signal morphology, dependent on the angle of transduction and the focus of the moving structure, created several problems.

4.4 Methods

In order to perceive the idea of the proposed method, consider training an HMM-based probabilistic model. Before applying the Viterbi algorithm, the emission probability density and duration distributions need to be established. Therefore, labeled data is needed as an input for the optimization methods such as the Baum-Welch algorithm to estimate mentioned distributions. However, in the absence of cardiac event labels, training the HMM parameters is challenging. Hence, unsupervised approaches can be used to avoid this limitation. We propose an unsupervised algorithm consisting of clustering for the recognition of states followed by probabilistic segmentation. An unsupervised learning approach groups data in a way such that similar objects, with respect to the feature set, will be labeled as the same group. By applying this on a per-signal basis, highly individual segmentation approaches can be created. Figure 4.2 shows the steps of the general approach.

In our proposed method, we assume that we have sequential data generated by switching dynamical systems, and each block of the data sequence is originated from a specific underlying distribution. Therefore, by tracking the changes in the observed distribution, we can detect the states of the signal. This method should reduce the effect of the changes in the pattern of the heartbeats since it detects the transitions in the signal dynamics. By considering the probability of being in each state as an input of an HSMM, the well-known dynamical programming approach of the Viterbi algorithm can be used to optimize the state sequence. An HSMM's strong statistical foundation will add to the robustness of the model and allow for the restriction of a physiologically plausible heartbeat duration.

4.4.1 Datasets

Two data sets were used in this work which were recorded in Leipzig University Hospital in Germany (dataset 1) and the Oxford John Radcliffe (JR) Hospital in the UK (dataset 2). Both data sets included 1D Doppler signals acquired using a hand-held device (AngelSounds Fetal Doppler JPD-100s, Jumper Medical Co., Ltd., Shenzhen, China) with an ultrasound transmission frequency of 3.3 MHz and a digitization sampling frequency of 44.1 kHz.

The dataset 1 consists of Doppler signals and simultaneously recorded seven channels of indirect abdominal ECG (1000 Hz) from five volunteers with pregnancies between the 20th and 27th week of gestation. Pathological cases such as IUGR, premature rupture of membranes, or fetal heart failure were also allowable inclusions for the recordings. The study was approved by the Leipzig University Hospital ethics committee (record 348-12-24092012) and written informed consent was obtained from each patient. Before the beat annotation, the DUS signals were downsampled to 4000 Hz using an anti-aliasing filter. Also, pre-processing algorithms including fECG denoising, fECG extraction and beat detection [29] were applied on fECG recordings using *FECGSYN* toolbox [18].

In the annotating procedure of dataset 1, two annotators corrected the location of the beats obtained from automatic fECG beat detection in addition to labeling the signal quality of each 3.75 seconds of data using the Matlab GUI. After excluding noisy segments, the data comprising 6521 beats was used for applying the segmentation algorithm. Figure 4.1 illustrates the opening and closing timings of the fetal aortic and mitral valves in relation to the simultaneous fECG.

As an independent test data, dataset 2 consists of 1D DUS signals from 17 healthy pregnant women who bore singletons between 20 and 38 weeks of gestation was used. The data were collected at the John Radcliffe Hospital in Oxford, UK. The study was approved by the NHS Health Research Authority (REC reference: 12/SC/0147),

Table 4.1: Summary of datasets used for developing 1D DUS segmentation. The Leipzig dataset was used for optimizing the parameters since it contains simultaneously recorded fECG, a validated reference technique used for fetal cardiac monitoring. The Oxford dataset was used as an independent test set.

Database	Subjects	Beats	Modalities	Beat annotation	Quality label	Use
Leipzig (Dataset 1)	5	6521	1D DUS, fECG, maternal ECG	correcting the location of the beats calculated from fECG	provided for 1D DUS and fECG channels	Model development
Oxford (Dataset 2)	17	4044	1D DUS	locating the beats by listening to 1D DUS	provided for 1D DUS	Model evaluation

and written consent was obtained from each study subject prior to data collection. For this data set labels of signal quality and location of the beats were provided by three trained annotators using a Matlab GUI. Annotators clicked for the beat sound to indicate the location of the beats and a correction algorithm was applied to the annotations as described in Valderrama *et al.* [252]. The 1D DUS recordings were re-sampled to 4000 Hz using an anti-aliasing filter, and noisy segments of the recordings were manually excluded. Dataset 2 comprising 4044 beats was used in the testing phase.

The details of the processing and annotating of dataset 1 and dataset 2 were provided in our previous work [252]. In this work, dataset 1 has been used to optimize the parameters and justify the capability of the method by cross-validated results. Dataset 2 was considered as an independent test data to evaluate the developed model. Table 4.1 provides the summary of information regarding dataset 1 and dataset 2.

4.4.2 Pre-processing

According to the method shown in figure 4.2, the first step in 1D DUS segmentation is pre-processing. Doppler signals, like other physiological observations, are often affected by internal and external interference such as respiration, movement, and environmental noise. To remove high-frequency noise and the baseline wander, a second-order band-pass Butterworth filter was applied on data. By observing the frequency components of the Doppler signals, the cut-off frequencies were set to 25

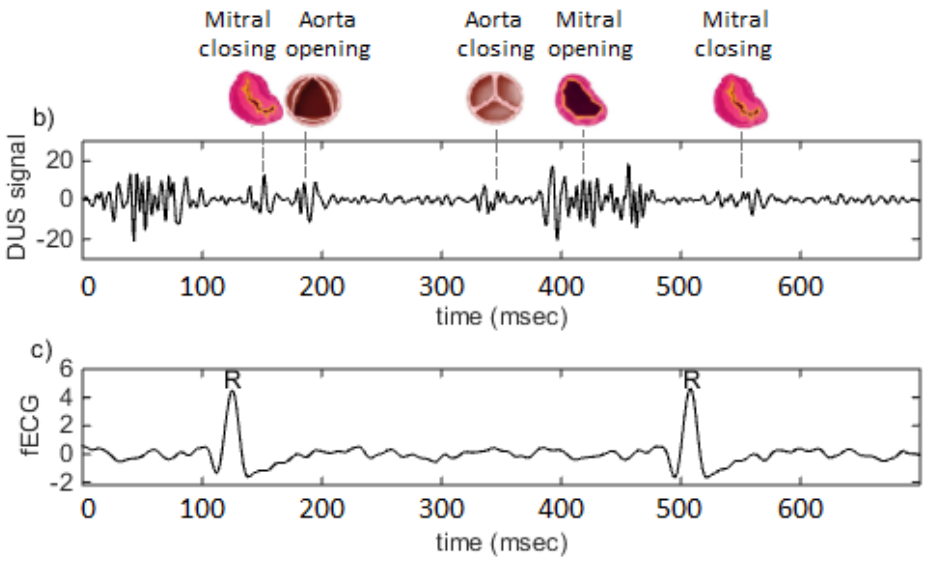


Figure 4.1: The opening and closing timings of the fetal aortic and mitral valves in relation to the fetal ECG: b) DUS signal. c) Fetal ECG recording. Adapted from [169].

to 600 Hz, corresponding to cardiac oscillations. Also, to remove transients or spikes, a spike removal filter was applied as per Schmidt *et al.* [215].

4.4.3 Data Transformation

The data transformation block in figure 4.2 includes feature extraction, time delay embedding and kernel density estimation (figure 4.2-b). The purpose of the data transformation block in this work is to reduce the variability relating to the highly variable DUS signal while keeping useful information in state transitions. The following sections specify the details of each data transformation step.

4.4.3.1 Feature Extraction

To characterize changes in time and frequency a set of features is established from DUS data (figure 4.2-b-I). We used Homomorphic envelope and power spectral density (PSD) to incorporate temporal and spectral properties of DUS data.

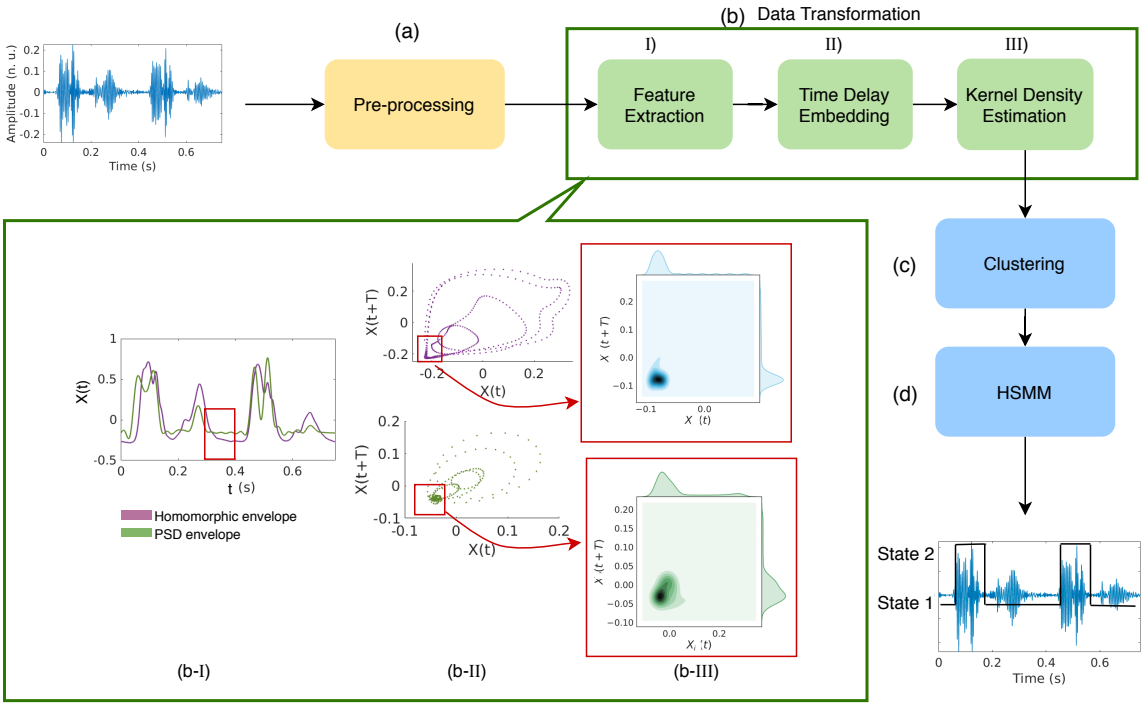


Figure 4.2: Illustration of the proposed unsupervised methodology for 1D DUS signal segmentation. a) The denoising filters are applied to 1D DUS. b) Filtered DUS signals are fed to the data transformation block, consisting of three steps: b-I) Feature extraction. b-II) Time delay embedding and, b-III) Kernel density estimation of windowed trajectories. The example of each step is also provided in the green box where (b-I) shows the Homomorphic and PSD envelopes, (b-II) indicates the time delay embedding of the two features and (b-III) is the kernel density of the specified window. c) The resulting kernels in overlapping windows are then clustered for recognizing the states of the data. d) Then, using the cluster centroids, the HSMM can estimate the most probable states of the data.

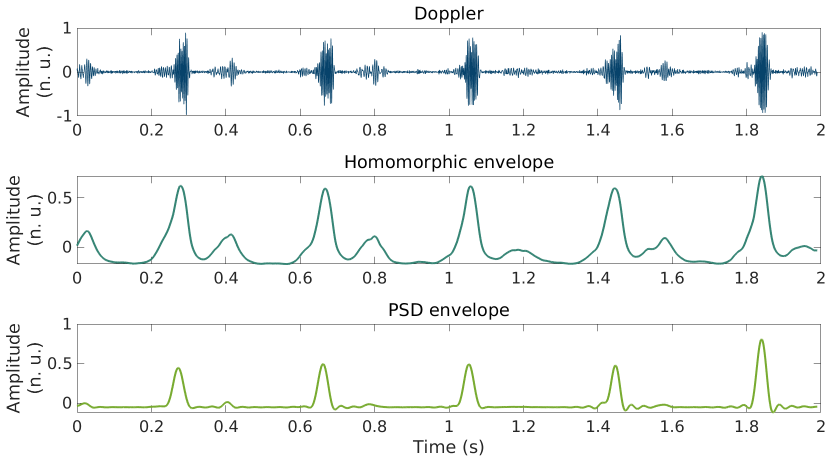


Figure 4.3: Extracted features from Doppler signal. From above figures are, the DUS signal, the Homomorphic envelope and the PSD envelope. © Institute of Physics and Engineering in Medicine. Reproduced with permission. All rights reserved (`https://iopscience.iop.org/article/10.1088/1361-6579/aba006/pdf`).

The Homomorphic envelope was used as a feature to capture changes in the amplitude of the DUS signal. This feature is suitable for the evaluation of signals combined by amplitude modulation or convolution [204]. To get Homomorphic envelope, the analytical signal was derived using the Hilbert transform. Then by applying a low-pass filter in the log-domain, the envelope of the analytical signal was separated.

The PSD envelope is used to extract the temporal and spectral information corresponding to cardiac valve and wall motion. The common time-frequency representation for non-stationary signals is short-time Fourier transform (STFT). STFT with Hamming window of 0.05 s in width with 50% overlap was used to calculate the PSD of a signal over time. For the data recorded with a 3.3 MHz frequency Doppler device, the cardiac movements generally take place within 100 Hz to 600 Hz. Therefore, the PSD envelope was derived by averaging over this range of frequency.

Figure 4.3 illustrates the example of Homomorphic and PSD envelopes used in this work. The feature vectors for each recording were individually normalised by subtracting their mean and dividing by their standard deviation. Following Springer *et al.* [228], to increase the speed of computation, the resulting feature vectors were

down-sampled to 500 Hz using a poly-phase anti-aliasing filter.

4.4.3.2 Time Delay Embedding

In this work, the time embedding method was used to represent features in a higher dimension and reveal the underlying dynamics of the features. An embedding space of features was formed by inclusion of temporal dependency (figure 4.2-b-II). Suppose that the observed measurement $x(t)$ is one dimensional data, the embedded data $\boldsymbol{z}_t = [x(t), x(t-\tau), ..., x(t-(m-1)\tau)]$ generally lies in some higher dimension m. Where m denotes the embedding dimension, and τ is a time delay parameter. In practice, reliable estimation of time delay τ and the minimum embedding dimension leads to a computationally effective model. In this work, an embedding dimension (m) was set to two and time delay τ was set to 10 ms using the false nearest neighbor method. As a result of embedding, for each feature, a periodic and recurrent behavior is characterized by a path returning to itself, creating one or more loops.

4.4.3.3 Kernel Density Estimation

The proposed method tracks the variation based on the estimated probability density functions (PDFs) in overlapping windows (figure 4.2-b-III). Considering that there is not an assumption on the underlying form of the distribution, a non-parametric method was employed to estimate the distribution directly from the data. The simplest non-parametric method to estimate the distribution is a histogram. However, the discontinuity in histograms in addition to the curse of dimensionality for high dimensional data makes it impractical. Also, for continuous data without a sufficiently large number of data points in a window, it is highly probable to see each data point once (uniform distribution). To estimate a PDF of the data points in each window, we used kernel density estimation. The formulation for the kernel density estimator

is

$$p_t(\boldsymbol{z}) = \frac{1}{N_w}\sum_{i=1}^{N_w}\frac{1}{(2\pi\sigma^2)^{m/2}}exp(-\frac{(\boldsymbol{z}-\boldsymbol{z}_i)^2}{2\sigma^2}), \tag{4.1}$$

where the kernel function was placed on each data point $\boldsymbol{z}_i$. The kernel density was estimated in overlapping windows of time-delayed data to track changes in the state of the signal. The window length (N_w) was optimized through cross-validation on dataset 1 and set to 80 ms. We determined a smoothing parameter (σ) by calculating the mean of distances of $\boldsymbol{z}_i$ from neighbors following Kohlmorgen *et al.* method [137].

4.4.4 Clustering

We leveraged a two-level clustering algorithm to group the PDFs and recognize the states in data (figure 4.2-c). A significant advantage of this clustering method is that it considers the time dependency of windows and it has a lower computational cost in comparison to clustering methods such as kmeans. The inputs of the proposed clustering algorithm are subsets of PDFs estimated from DUS 3.75 s segments. The duration of 3.75 s was fixed since it is the usual length based on the Dawes/Redman criteria [192, 62] and we expect the higher cluster purity in each 3.75 s segment of DUS.

For the first-level clustering, a time-dependent clustering method was used. The time dependency was considered by clustering the data incrementally and updating the clusters in each step by observing a new PDF. Briefly, in this approach, an upcoming PDF which is not similar enough to the existing clusters will create a new cluster. The threshold value for creating a new cluster was optimized based on the desired number of states.

The second level of the clustering algorithm was then performed by applying a k-means clustering method on cluster centroids calculated from the first level. Therefore, the final cluster centroids were calculated by applying the second level clustering.

Cluster centroids were then considered as the state representative. In both clustering levels, energy distance measure was used which is a statistical distance between the distributions. Following equation represents the distance measure between two PDFs p_i and p_j:

$$d(p_i(\boldsymbol{z}), p_j(\boldsymbol{z})) = \int (p_i(\boldsymbol{z}) - p_j(\boldsymbol{z}))^2 d\boldsymbol{z}. \tag{4.2}$$

Defining the states of DUS signal segmentation model is challenging due to the changes in morphology of this signal and an uncertainty about which valve or wall motion will be dominant in each cycle. In this study the number of states is set to two, following the model presented in [234]. Therefore, states of the proposed model indicates the dominant beat event in each cycle and the intervals between them. For each upcoming window, the probability of being in each state can be calculated based on the distance to the cluster centroids.

4.4.5 Hidden Semi-Markov Model

HMM is a powerful statistical model for the segmentation of sequential observation. The underlying assumption of the statistical models is that the signal is characterized as a parametric random process and the parameters can be estimated in a well-defined manner [199]. The major difference between HSMM and HMM is the assumption that the underlying stochastic process is a "semi-Markov" process. In this approach, the duration of each state is considered as a random variable characterized by an explicit probability distribution. As a result of this modification, self-transition coefficients in the transition matrix are set to zero. The duration densities control the appropriate amount of observations required for each state. In this work, the estimation of HSMM variables was modified to use it as an unsupervised model. To estimate the most probable state sequence $Q = \{q_1 q_2 ... q_T\}$ for a given observation sequence $O = \{O_1 O_2 ... O_T\}$ the well-known dynamic programming algorithm, Viterbi algorithm

is used. The likelihood of the most probable state sequence corresponding to the first t observations and ends in state i can be expressed as:

$$\delta_i(t) = \max_{q_1 q_2 \cdots q_{t-1}} p(q_1 q_2 ... q_t = i, O_1 O_2 ... O_t | \lambda), \tag{4.3}$$

in which λ denotes the parameters of the model. Generally, an HMM can be characterized by its parameters as:

$$\lambda = (A, B, \pi), \tag{4.4}$$

where A is the transition probability matrix, B is the emission matrix and π is the initial state distribution. The size of the model parameters is equal to the number of states recognized in the clustering algorithm where we can define the hidden states as $\mathbb{S} = \{s_1, s_2, ..., s_N\}$. In this work the total number of states N is set to two.

4.4.5.1 The initial state distribution:

The initial state probability $\pi = \{\pi_i\}$ was defined as a uniform distribution since there was not any restriction for assigning a label to the first observation and it is equally likely to observe either state.

4.4.5.2 The Observation Probability Density

The observation probability density $B = \{b_j(O_t)\}$ defines the probability that state j generates the observation vector O_t at time t and was deployed based on the similarity of the PDFs to the cluster centers. In this work, the probability of observing the estimated PDF in state $s \in \mathbb{S}$ is defined as:

$$b_j(p_t(\boldsymbol{z})) = p(p_t(\boldsymbol{z})|s_j) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{d(p_{s_j}(\boldsymbol{z}), p_t(\boldsymbol{z}))}{2\sigma^2}), \tag{4.5}$$

in which $p_s(\boldsymbol{z})$ is the cluster centroid and $p_t(\boldsymbol{z})$ is the estimated PDF from observation.

4.4.5.3 The duration distribution of Doppler states

We used a Gaussian probability density function for the duration probabilities, which can capture the inherent variation of beat duration. The assumption in pruning theorem requires the state duration distributions to be log-convex, as is the case for most parametric distributions useful for the purpose [285].

The parameters of the duration distributions ($p_i(d)$) were initialized based on physiological limitations, then we updated the distributions based on the estimated mean of cardiac cycle duration (μ). The mean cardiac cycle duration in each 3.75 s of DUS was computed by auto-correlation of the signal envelope using the method presented by Valderrama *et al.* [252]. The mean cardiac cycle also was used to specify the maximum duration (d_{max}) in the Viterbi algorithm.

Note that, estimating the duration distribution of each state is required for the HSMM. However, the auto-correlation method provides the estimate of the mean beat interval which is the summation of the duration of two states ($\mu = \mu_{s_1} + \mu_{s_2}$). We proposed an adaptive algorithm to update the initial distributions based on estimated mean heart rate using Kullback-Leibler (KL) divergence. Specifically, KL divergence of reference density P and estimated density Q is a measure of the information lost when P is used to approximate the reference. In this work, P is the initial duration distribution and Q is the updated version of P using mean of the duration of the cardiac cycles. The formulation for KL divergence is as follow:

$$L = KL(Q\|P) = \sum Qln\frac{Q}{P}. \tag{4.6}$$

The optimization process is based on one variable since, μ_{s_1} and μ_{s_2} are dependent.

The defined objective function is

$$L_1 = L(N(\mu_{s'_1}, \sigma_{s_1}) \| N(\mu_{s_1}, \sigma_{s_1})), \tag{4.7}$$

$$L_2 = L(N(\mu - \mu_{s'_1}, \sigma_{s_2}) \| N(\mu_{s_2}, \sigma_{s_2})), \tag{4.8}$$

$$\underset{\mu_{s'_1}}{\operatorname{argmin}}(L_1 + L_2), \tag{4.9}$$

where μ_{s_1} and μ_{s_2} are the initial values and $\mu_{s'_1}$ is the updated value of mean for state one and variable of the objective function.

Heretofore, we discussed the definition of parameters in our HSMM model. Ultimately, the Viterbi algorithm was used to estimate the most likely sequence of PDFs that might have generated the given sequence of PDFs. The likelihood of the most probable sequence can be estimated using the following recursive relation:

$$\delta_i(t) = \max_d \{\max_j \{\delta_{t-d}(j) a_{ij}\} p_i(d) \Pi_{t'=t-d+1}^{t} b_i(p_{t'})\}, \tag{4.10}$$

where the maximization was performed over all the defined states and time samples $1 \leq i, j < N$, $1 \leq t \leq T$ and all possible state durations d where $1 \leq d \leq d_{max}$.

4.5 Performance assessment

We investigated the effect of the window size used in kernel density estimation on $F1$ score of beat detection and mean square error (RMSE) in capturing variability using dataset 1. For this purpose, the simultaneous fECG was considered as a reference for benchmarking. The leave-one-out cross-validation method was used as a model validation and parameter tuning technique to divide the data set into the training and validation sets.

After parameter optimization, two approaches were used to reach a consensus on

the performance of the algorithm. First, the standard evaluation metrics, sensitivity (Se), positive predictive value (PPV), and $F1$ score were provided to measure the beat detection accuracy. Also, the RMSE of the variability in detected beat intervals was calculated. The second aspect of the evaluation is the performance of the algorithm in HRV estimation. Therefore, the accuracy of the beat to beat interval estimation was investigated by comparing the obtained HRV metrics from DUS and fECG in frequency and time domain. HRV analysis included the mean of intervals (NNmean), the standard deviation (SDNN), the root mean square of successive differences (RMSSD), the proportion of interval differences of successive intervals greater than 20 ms (pNN20), the low frequency power (LF), and the high frequency power (HF). In this work, HRV was analyzed using a 3-minutes sliding window with the increment of 30 seconds. Based on the literature report on FHRV spectral analysis in [205] and the successful use of LF and HF range to discriminate abnormal cases in [15], the frequency bands of interest for LF and HF were set to 0.04-0.15 and 0.15-0.5 Hz, respectively. An open-source "*PhysioNet Cardiovascular Signal*" toolbox was used to calculate the HRV indices [255].

The statistical comparison tests are provided to analyze the HRV estimation. The Spearman's rank-order correlation and the corresponding p-values tests the null hypothesis that there is no relationship between the variables (α=0.05). In addition, the p-value of a two-sided Wilcoxon rank-sum test was considered as a second measure to test the null hypothesis that inputs are samples from continuous distributions with equal medians against the alternative (α=0.05).

Figure 4.4 shows the two-state segmentation result. The performance of the algorithm was evaluated based on the transition point between the states. Therefore, three fiducial points were extracted from segmentation labels, including onset, middle, and offset of the dominant beat. The intervals between the mentioned points are denoted by Δ_1^D, Δ_2^D and Δ_3^D respectively. Note that the fiducial point derived from

HSMM may not necessarily identify the same location as an annotation. However, the number of the detected beats and variability of the intervals are important to evaluate the performance.

4.6 Results

4.6.1 Analysis of dataset 1

The effect of the window length was investigated using dataset 1. The $F1$ score and RMSE values are provided in Figure 4.5. An extremely small RMSE shows that the detected beats were located accurately. Considering the computation cost for increasing the window size and the results in Figure 4.5, we can infer that choosing the second window size (80 ms) leads to having good performance in beat detection and more accurate overall variability estimation.

The segmentation algorithm was applied based on the optimized window size. Table 4.2 includes the beat detection accuracy of dataset 1 averaged across the subjects to an overall accuracy of 97.5%, 97.9% and 98.2% F1 score for Δ_1^D , Δ_2^D and Δ_3^D respectively. By using Δ_3^D, better performance in beat detection was achieved, and the reason is the essence of the annotation in dataset 1, which is based on fQRS location.

The accuracy of estimated beat-to-beat intervals is provided for each patient in terms of HRV analysis (table 4.3). Overall, results show that the onset-to-onset interval (Δ_1^D) is the best estimate for capturing the heart rate variability. The estimated FHRV parameters and the reference values have significant correlation in terms of Spearman correlation metric. Also, based on the Wilcoxon rank sum test, we can infer that there is insufficient evidence to reject the null hypothesis (sampling from continuous distributions with equal medians) in the estimates of the Mean, SDNN and LF metrics calculated from dataset 1. Figure 4.6 shows the violin plot of the

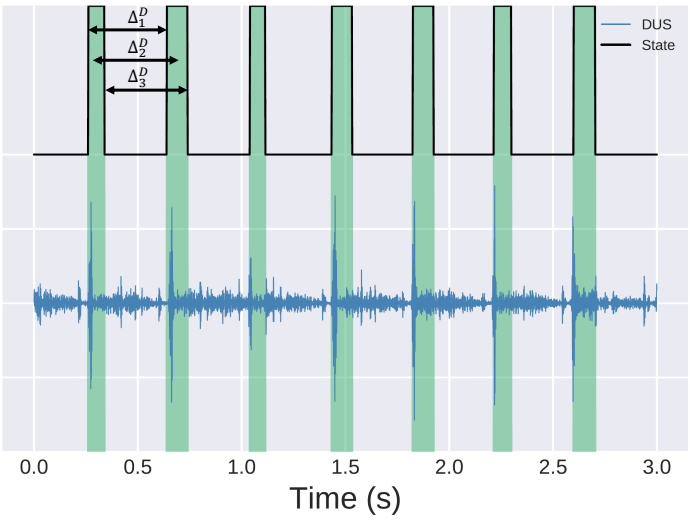


Figure 4.4: The segmentation result and simultaneous fECG from dataset 1. The performance of the method was evaluated using start, middle and end of the dominant beat. The estimated intervals were denoted by Δ_1^D, Δ_2^D and Δ_3^D respectively.

Table 4.2: The performance of proposed unsupervised HSMM beat-to-beat interval estimation on dataset 1.

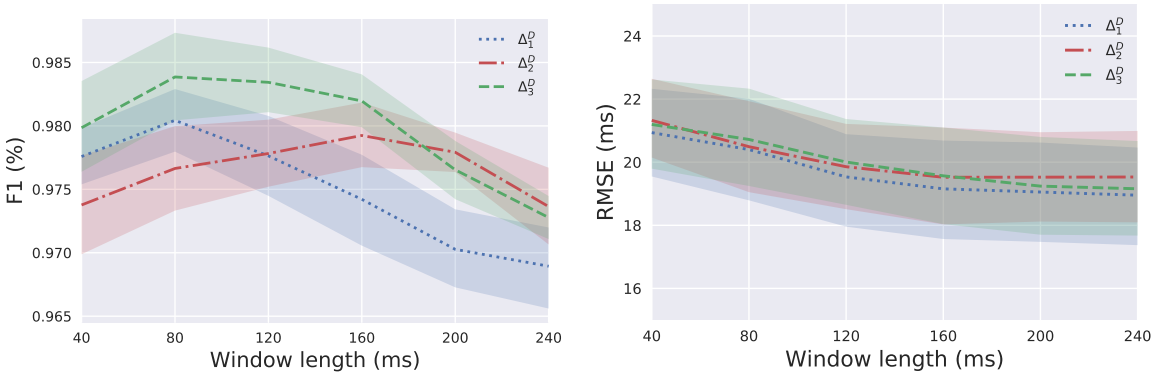
Se (%)			PPV (%)			F1 (%)			RMSE (ms)		
Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D
97.6	97.6	98.4	97.3	98.0	98.1	97.5	97.9	98.2	20.40	20.72	20.49

HRV metrics for all the patients in dataset 1, which demonstrates the distribution of the HRV indices derived from the estimated and reference values.

4.6.2 Analysis of dataset 2

As we described in the study database, the annotation for dataset 2 is not based on fECG and provided by three independent annotators. This data set is used as an independent test data to evaluate the performance of the algorithm using calculated cluster centroids and optimized window size.

Table 4.4 shows a performance of the algorithm on dataset 2 with F1 score ranged



(a) (b)

Figure 4.5: The cross-validated measures to optimize window length. a) F1 score of beat detection and b) RMSE of variability estimation for a range of window sizes. © Institute of Physics and Engineering in Medicine. Reproduced with permission. All rights reserved (`https://iopscience.iop.org/article/10.1088/1361-6579/aba006/pdf`).

Table 4.3: The performance of unsupervised HSMM heart rate variability estimation on dataset 1. ρ_{re} is the correlation coefficient between the reference and estimated HRV metrics, P_S denotes the p-value for the Spearman correlation and P_W is a p-value of the Wilcoxon rank sum test with the significance level of $P < 0.05$.

	ρ_{re}			P_S			P_W		
	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D
NNmean	0.99	0.99	0.99	0	0	0	0.22	0.24	0.24
SDNN	0.92	0.91	0.91	0	0	0	0.7	0.4	0.6
RMSSD	0.56	0.50	0.47	1.1e-04	9.5e-04	6.5e-04	3.5e-06	2.5e-06	5.7e-06
pNN20	0.50	0.48	0.42	9.1e-04	0.004	5.1e-04	1.1e-06	5.2e-06	4.8e-06
LF	0.98	0.98	0.98	0	0	0	0.72	0.88	0.74
HF	0.60	0.43	0.30	0.01	0.01	0.04	0.004	0.004	0.004

from 97.7% to 98.5%. Based on the results provided in table 4.5 and Figure 4.7, we can conclude that all the estimated FHRV metrics can follow the reference values in terms of Spearman correlation. The Wilcoxon rank sum test indicates the rejection of null hypothesis (no statistically significant difference in medians) for the pNN20 measure. In addition, by comparing the three estimated intervals Δ_1^D, Δ_2^D and Δ_3^D, we can infer that Δ_1^D is the most accurate estimation in terms of HRV assessment and beat detection.

Table 4.4: The performance of proposed unsupervised HSMM beat-to-beat interval estimation on dataset 2.

Se (%)			PPV (%)			F1 (%)			RMSE (ms)		
Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D
98.0	97.2	97.2	98.9	98.9	98.1	98.5	98.0	97.7	26.3	27.1	27.4

Table 4.5: The performance of unsupervised HSMM heart rate variability estimation on dataset 2. ρ_{re} is the correlation coefficient between the reference and estimated HRV metrics, P_S denotes the p-value for the Spearman correlation and P_W is a p-value of the Wilcoxon rank sum test with the significance level of $P < 0.05$.

	ρ_{re}			P_S			P_W		
	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D	Δ_1^D	Δ_2^D	Δ_3^D
NNmean	0.99	0.99	0.99	0	0	0	0.26	0.26	0.25
SDNN	0.92	0.89	0.87	0	0	0	0.06	0.04	0.003
RMSSD	0.67	0.62	0.67	1.6e-07	1.5e-06	1.5e-07	0.78	0.78	0.41
pNN20	0.65	0.63	0.57	1.0e-08	5.5e-07	8.7e-06	0.001	0.002	0.001
LF	0.92	0.91	0.94	0	0	0	0.09	0.05	0.01
HF	0.65	0.52	0.33	3.4e-07	9.5e-05	0.01	0.41	0.33	0.20

4.7 Discussion

In this work, we have presented a novel approach to signal segmentation using an unsupervised HSMM. Previous methods of FHR estimation using 1D DUS focused on two alternative approaches with associated weaknesses. The first approach is supervised and requires labeled data to train a segmentation model [173, 172] and is therefore dependent on the quality of the labels, and the similarity between the labeled data and the new data encountered. The second approach ignores beat events and involves FHR estimation using fixed temporal windows of DUS data. The most popular approach in this latter category is based on an auto-correlation method [252] which provides a lower resolution for heart rate timing estimation than segmentation approaches due to averaging over a few seconds. The EMD-Kurtosis approach of [3], which also falls into the non-beat level windowed FHR estimation category, has been shown to provide more accurate estimations than standard auto-correlation based

methods. The reported results showed that although the EMD-Kurtosis method could outperform the auto-correlation approach in estimating the mean beat-to-beat interval, there were statistically significant differences in the derived estimates of the SDNN and RMSSD metrics. This indicates that the EMD-Kurtosis approach introduced errors in HRV estimation for all metrics evaluated.

The statistical analysis of our proposed method in terms of FHR variability metrics (table 4.3 and table 4.5) substantiates our method's superiority to this other work. Specifically, we see no statistically significant differences in our estimates of all FHRV metrics in terms of Spearman correlation in dataset 1 and dataset 2. Moreover, the Wilcoxon rank-sum test rejects the null hypothesis of equal medians in estimates for the RMSSD, pNN20 and HF HRV metrics calculated from dataset 1. For dataset 2 the Wilcoxon rank-sum test indicates the rejection of the null hypothesis only for the pNN20 metric. All other metrics cannot be shown to be different to the human annotations. Finally, we note that in our work (unlike previous works) we used an independent test dataset drawn from a different population and setting, and so the generalization of our reported performance is likely to be more realistic.

Based on the results shown in table 4.3 and table 4.5, some FHR variability metrics such as RMSSD, pNN20, and HF are more prone to errors than others depending on the selected fiducial point (Δ_1^D, Δ_2^D or Δ_3^D). The sensitivity of pNN20 and RMSSD to the fiducial point choice can be explained considering their definition, which is based on successive differences. Therefore, the possible error in the beat location may influence these parameters more than the SDNN. Also, we note that the HF estimate has been shown to be very sensitive to small changes in beat onset detection [55].

It should be noted that the heart rate estimations are affected when data are heavily corrupted by noise and interference [252], and in fact it is important at that point to not report heart rate, or derived HRV metrics, but rather to report that the data

are non-analyzable. In our earlier works [236, 250, 251], we demonstrated accurate methods for separating poor quality from good quality data, and even identifying the etiology of the noise. Therefore, coupling these works together may lead to a robust system that could be used in an automatic or semi-automatic manner. We also note that it was impossible for the human annotators to accurately estimate beat onsets and fetal heart rate from noisy DUS data, and therefore assessments during noisy sections are not possible.

Comparisons of the results for detecting the three possible fiducial points in DUS (onset, offset and the midpoint of the beat) demonstrates that the interval estimated from the onset of the dominant beat (Δ_1^D) is the most robust measurement of beat interval for assessing HRV. This could be due to the nature of the proposed method, which detects changes in the dynamics of the observation based on the learned parameters and the offset and midpoint can sometimes exhibit more indefinite fiducial points. Furthermore, the onset time might best reflect the initiation of the cardiac cycle and be less confounded by other extraneous factors.

The proposed method alleviates the cost of labeling clinical data by using unsupervised learning to estimate beat-to-beat intervals, which has been shown to be important in various applications, including evaluation of fetal maturation [111] and comparison of HRV indices of low-risk and IUGR fetuses [229, 217, 15]. In particular, SDNN, RMSSD, HF, and LF have shown to be significantly lower in IUGR fetuses. Since our approach is capable of estimating SDNN and LF with minimal error (correlation coefficient>0.9) the developed framework can be thought of as a low-cost promising tool for FHR monitoring. In particular, the proposed technique may have potential for improving the clinical utility of IUGR classification and prediction from FHRV [238]. However, future work prospectively analyzing a significant volume of IUGR fetuses with raw Doppler will be required to prove the eventual utility of such an approach.

An inevitable limitation in the processing of the clinical data is the existence of noise. In DUS recording sources of noise include fetal movement, probe movement or disconnection, maternal movement, and other moving organs and environmental noise. The existence of low-quality segments of DUS recording affects the estimation of most HRV metrics. Another limitation of this study is that the reference signal is a non-invasive fECG, which is sometimes noisy and prone to errors at times [208]. A scalp fECG could have provided a more accurate and more reliable reference for evaluation, but it requires the rupture of membranes and can only be performed during labour.

4.8 Conclusion

In summary, this work presents the most accurate approach for the beat to beat monitoring of fetuses using 1D DUS signals so far reported. The proposed modified HSMM as an unsupervised approach along with the use of spectral and temporal features as an input to a clustering algorithm, significantly improved DUS segmentation and heart rate variability assessment. This method may provide improved FHR monitoring from low-cost 1D DUS transducers, and assist in better detection of fetal abnormalities in low-resource settings.

Acknowledgments

GC and NK acknowledge the support of the National Institutes of Health, the Fogarty International Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number 1R21HD084114-01 (Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala). LS acknowledges the support of the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation) and

of the Oxford Centre for Affordable Healthcare Technology. CV is funded by a Fulbright Scholarship. The authors acknowledge the collaboration of the Institute of Biomedical Engineering, TU Dresden, and the Department of Obstetrics, University Hospital of Leipzig, for providing Doppler ultrasound and fetal electrocardiogram recordings.

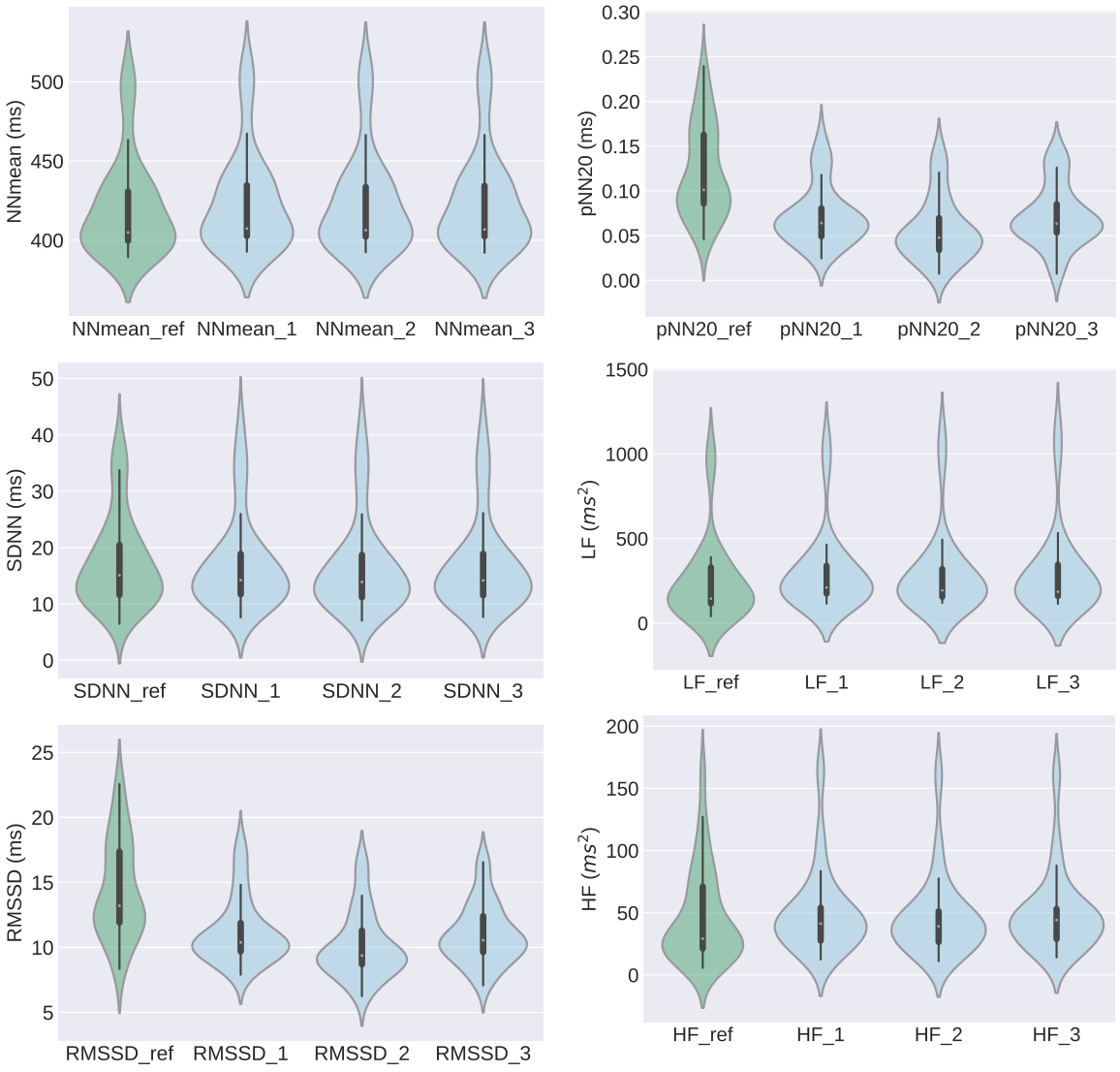


Figure 4.6: Comparison of estimated fetal HRV parameters from DUS segmentation and reference values (beat intervals from fECG) in dataset 1.

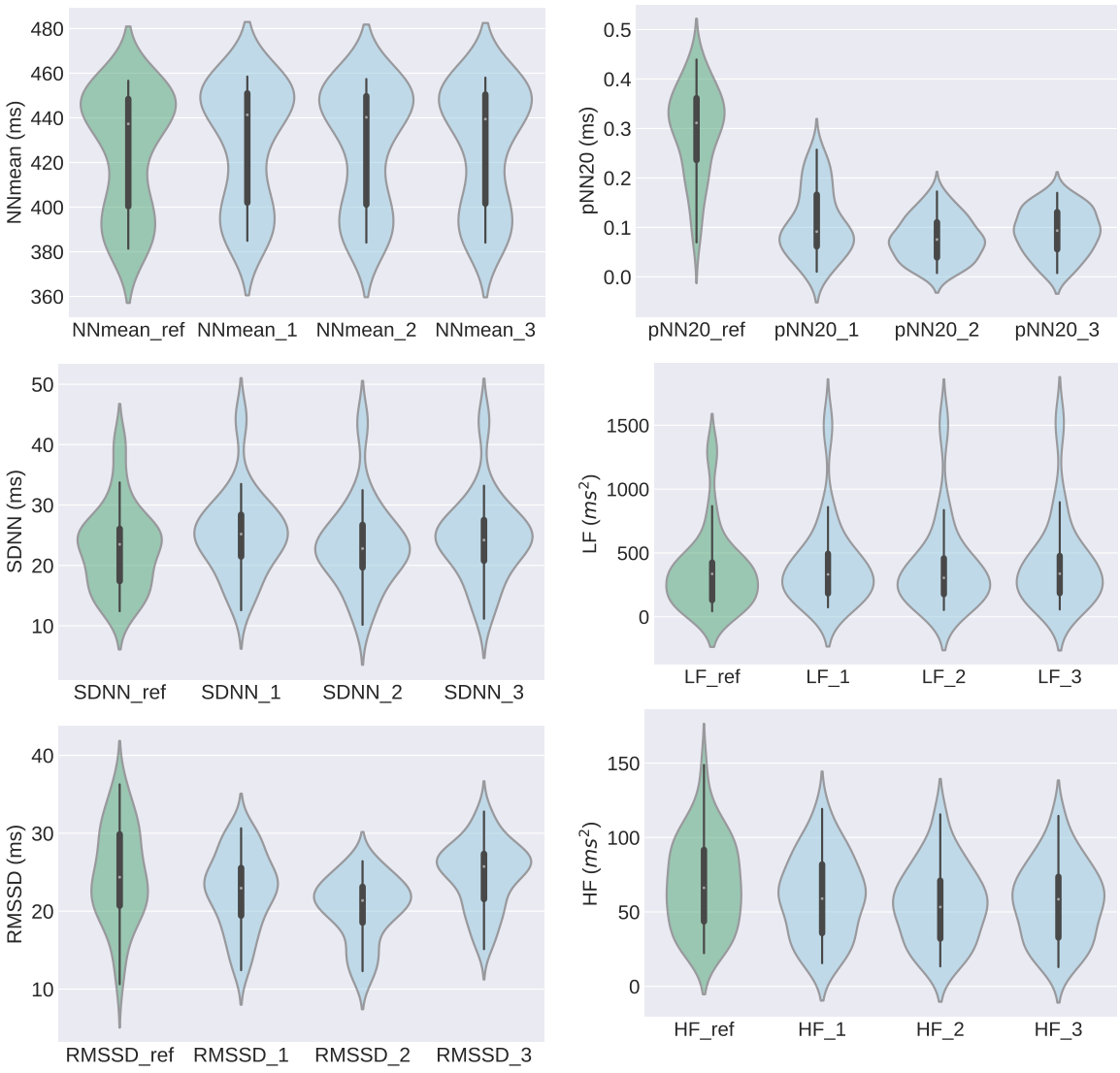


Figure 4.7: Comparison of estimated fetal HRV parameters from DUS segmentation and reference values (beat annotation) in dataset 2. © Institute of Physics and Engineering in Medicine. Reproduced with permission. All rights reserved (`https://iopscience.iop.org/article/10.1088/1361-6579/aba006/pdf`).

Chapter 5

Deep sequence learning for gestational age estimation

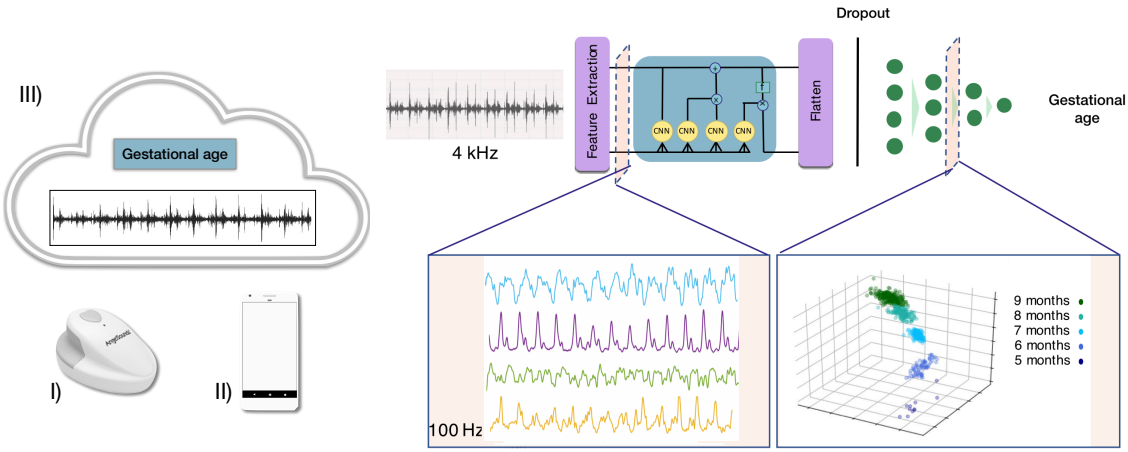
5.1 Abstract

Assessing fetal development is usually carried out by techniques such as ultrasound imaging, which is generally unavailable in rural areas due to the high cost, maintenance, skills and training needed to operate the devices effectively. In this work, we propose a low-cost one-dimensional Doppler-based method for estimating gestational age (GA). Doppler time series were collected from 401 pregnancies between 5 and 9 months GA using a smartphone. The proposed model for GA estimation is based on sequence learning by forming a temporally dependent model using a convolutional long-short-term memory network. Time-frequency features are extracted from Doppler signals and regularized before feeding to the network. The overall mean absolute GA error with respect to the last menstrual period was found to be 0.71 month, which outperforms all previous works.

5.2 Introduction

Low- and middle-income countries (LMICs) account for approximately 98% of all reported perinatal deaths worldwide, mainly due to gestational developmental issues, specifically intrauterine growth restriction (IUGR) [286, 151, 59]. Rising costs of healthcare and inadequate access to prenatal medical services exacerbate this issue in LMICs as well as low-income regions in other countries, such as the southeast US. Most of these deaths can be avoided by improving health monitoring before, during and after childbirth. Therefore, developing AI-enabled edge-computing devices that are intuitive to use, even for low-literacy populations helps to enhance healthcare for disadvantaged populations.

Gestational age estimation provides essential information such as preterm birth management, delivery scheduling and growth restriction [5]. Fetal cardiac assessment is a tool recommended by obstetrical societies for monitoring fetal health during pregnancy [154]. The functional assessment of the fetal heart conveys important information regarding the hemodynamic status and cardiovascular adaptation of a fetus in the face of several perinatal complications. Fetal heart rate is influenced by the autonomic nervous system (ANS), which matures during pregnancy. In particular, fetal heart rate variability evolves over the course of pregnancy reflecting the maturity of the ANS, and thus an indirect indicator of the fetal gestational age [258]. Previous studies have shown that fetal heart rate variability metrics can be used as discriminative features for fetal development assessment [248, 110, 240, 170, 171]. One non-invasive method for capturing fetal cardiac activity is one-dimensional Doppler ultrasound (1D-DUS), which is a low-cost and simple method for fetal heart rate monitoring [174]. The Doppler transducer can be easily adapted to connect to mobile devices such as smartphones, for recording and processing, motivating their use in mobile-health (mhealth) systems for risk screening in low-resource environments [237]. The Doppler transducer transmits and receives ultrasound waves, which reflect fetal



(a) Data collection (b) Gestational age estimation model

Figure 5.1: (a) Data collection. I) The raw 1D ultrasound is captured using Doppler transducer and II) The 1D-DUS and gestational age are recorded on the phone using the developed mobile app. III) The data are then uploaded to the cloud for backup and further processing. (b) An overview of the proposed process for training CLSTM network for fetal monitoring from abdominal Doppler acquired during routine fetal monitoring. The features from 1D-DUS are calculated and fed to the CLSTM network. The output is then flattened and mapped to the target label.

cardiac activity. Using 1D-DUS signal, blood flow, cardiac wall, and valve motions can be captured and are differentiable based on their different velocities. Although 1D-DUS provides useful information regarding cardiac functionality, its variable morphology makes the processing and modeling of this signal challenging.

In this paper, we propose a systematic approach for assessing fetal development by discovering the relation between fetal 1D-DUS signal and gestational age. The proposed approach is based on a set of effective time-frequency domain features of the 1D-DUS and a convolutional long-short-term memory (CLSTM) network [281], which is a robust and powerful method for extracting features from sequential data. This approach is used to model time dependencies in fetal 1D-DUS and to capture the variability of the cardiac activity, eventually leading to the estimation of the fetal gestational age. In the sequel, we start by formulating a mathematical model for the problem of fetal gestational age estimation from 1D-DUS signals.

5.3 Gestational age estimation model

Let $x(n;t)$ denote the time series of a 1D-DUS signal with discrete time index n, acquired during a clinical visit of a pregnant woman on date t. For simplicity, the date and gestational age are represented in units of weeks. The "true gestational age" at date t is denoted $a(t) = t - c$, where c is the *date of conception*, while the presumed or reported gestational age is $\tilde{a}(t) = t - \tilde{c}$, where $\tilde{c}$ is the *anticipated conception date*. Therefore, the presumed (anticipated) and true gestational ages can be related as follows:

$$\tilde{a}(t) = a(t) + \eta \tag{5.1}$$

where $\eta = c - \tilde{c}$ is the gestational age presumption error, which without additional priors (such as 2D-Doppler) remains an unknown stochastic constant over pregnancy. The error η accounts for lack of knowledge of the last menstrual period and uncertainties in the exact ovulation, intercourse and conception dates. We further denote the p-dimensional feature vector extracted from the 1D-DUS by $\mathbf{f}(x(n;t)) \in \mathbb{R}^p$. The objective is to design a deep network that estimated the true gestational age from the feature vector extracted from a single or a set of 1D-DUS acquired during pregnancy, i.e.,

$$\hat{a}(t) = G\left[\tilde{a}(t), \{\mathbf{f}(x(n;t_k))\}_{k=1}^{L}\right] \tag{5.2}$$

where $\hat{a}(t)$ is an estimate of the true gestational age, t_k $(k = 1, \ldots, L)$ denote the L dates that 1D-DUS is acquired from the pregnant woman, and $G(\cdot)$ denotes the feature-vector to gestational age transform that is learned by the neural network, as shown in Fig. 5.1-(b). In this scheme, the presumed gestational age $\tilde{a}(t)$ is used for model training.

5.4 Dataset

A hand-held 1D-DUS device, the Angel Sounds Fetal 1D-DUS JPD-100s (Jumper Medical Co., Ltd., Shenzhen, China), with an ultrasound transmission frequency of 3.3 MHz, and costing $25, was used to capture audio data from 401 pregnant women (493 visits and 693 recordings) at 5 to 9 months of gestation. The data, collected as part of a randomized control trial conducted in rural highland Guatemala [165, 251], include 15, 77, 162, 186, 253 recordings corresponding to gestational ages of 5, 6, 7, 8 and 9 months, respectively. The 1D-DUS signals were recorded by traditional birth attendants, who were trained to use the hand-held 1D-DUS device and were an accompanying mobile application. Immediately before recording the 1D-DUS signals, the traditional birth attendants entered the estimated gestational age into the app in months, based on the last menstrual period (LMP). Data were captured using a bespoke Android client at 44.1 kHz, using a low-cost smartphone (Samsung S3 mini) and stored as uncompressed WAV files at 7056/s bits) [237]. The first five minutes of each recording were used. Figure 5.1-(a) illustrates the data sources and devices used in this research.

5.5 Data analysis and feature extraction

5.5.1 Preprocessing

Given the nature of the physiological time-series, 1D-DUS signals are corrupted with internal and external interference such as respiration, movement, and environmental noise. In this work, a second-order band-pass Butterworth filter was used to reduce the noise. By observing the frequency components of the 1D-DUS signals, the cut-off frequencies were set to 25 and 600 Hz, corresponding to cardiac oscillations.

5.5.2 Time-frequency (TF) features for DUS components

For a real-valued discrete-time signal x_n, where n is the time instant, we define a windowed version of the signal $s_n = w_n x_n$, where w_m $(m = 0, \ldots, N-1)$ is a window for improving the spectral features and minimizing the windowing effects. A Hamming window of length 100 ms (400 samples at a sampling rate of 4 kHz) is used for the later presented results. The discrete-time Fourier transform (DTFT) of a window of N samples of s_n is:

$$S_n(\omega) \triangleq \sum_{m=0}^{N-1} s_{n+m-N+1} e^{-j\omega m} \tag{5.3}$$

According to the Parseval's theorem, the energy of each window of the signal is:

$$E_n \triangleq \sum_{m=0}^{N-1} |s_{n-m}|^2 = \frac{1}{\pi} \int_0^{\pi} |S_n(\omega)|^2 d\omega \tag{5.4}$$

We define the *instantaneous frequency*, or the *first spectral moment* of s_n, as follows:

$$\omega_n \triangleq \frac{1}{E_n} \int_0^{\pi} \omega |S_n(\omega)|^2 d\omega \tag{5.5}$$

which for frequency-domain unimodal signals is the frequency (in radians) around which the signal energy is localized at time instant n. It is therefore a measure of the signal's center frequency at time n. In a similar manner, the *instantaneous bandwidth*, or the *spectral centralized second moment*, of s_n is defined:

$$\Delta\omega_n^2 \triangleq \frac{1}{E_n} \int_0^{\pi} (\omega - \omega_n)^2 |S_n(\omega)|^2 d\omega \tag{5.6}$$

which is a measure of the instantaneous energy spread around the instantaneous frequency.

Finally, we define the instantaneous oscillation quality-factor (*Q-factor*): $Q_n \triangleq \omega_n / \Delta\omega_n$, as a measure of oscillation quality, which is a notion commonly used in elec-

tronic circuitry for evaluating the quality of oscillation independent of the frequency. Accordingly, when x_n becomes closer to a single-tone component, the Q-factor increases. Note that for digital implementations, the DTFT is replaced by the Discrete Fourier Transform (DFT), with appropriate dimension corrections. These features form the overall feature vector $\mathbf{f}_n \triangleq (\sqrt{E_n}, \omega_n, \Delta\omega_n^2, Q_n)$, which is fed to the sequence modeling part of the model. Due to the signal windowing, the extracted feature vector has slow variations over time and was therefore resampled from 4 kHz to 100 Hz, to reduce the processing load.

5.5.3 Sequence modeling

Given that we have a sequential feature vector, the use of a recurrent neural network is a natural choice to keep track of the variability and temporal structure of the signal. Long-short-term memory (LSTM) [103] is one such recurrent neural network, which has been used in various studies for the general purpose of sequence modeling. A Convolutional LSTM (CLSTM) network developed by Shi *et al.* [281] is a combination of LSTM and convolutional neural network to capture spatio-temporal features. Accordingly, the input-to-state and state-to-state transitions in LSTM are changed from full connections to convolution structure. By stacking multiple CLSTM layers, one can form a spatio-temporal sequence modeling network to uncover the variability in fetal cardiac activity. By changing the kernel size, CLSTM is able to capture the different DUS components with different velocities, corresponding to the different fetal-maternal body tissues that move within the DUS transceiver frequency range. We can consider the states as the hidden representation of the cardiac valve opening and closing. These sub-organ motions are captured by setting the appropriate kernel size. The utilized CLSTM architecture comprised of two successive layers with kernel sizes $(1, 20)$ and $(1, 10)$, respectively. Each layer is followed by batch normalization. The output is then flattened and mapped to the gestational age label through 4 fully

connected layers with sizes 128, 32, 3 and 1. In order to reduce the likelihood of overfitting we used an L2-regularizer with regularization parameter $\lambda = 0.01$. A dropout technique was also used before the dense layer and the probability of training a given node in a layer was set to 0.3.

5.6 Results

Stratified five-fold cross-validation is used across patients to assess the performance of gestational age estimation. The model was trained end-to-end for a total of 300 epochs using a mean absolute error (MAE) loss function. The batch size was fixed to 32 patients and generated using a balanced batch generator with random oversampling (with replacement) of the less frequent label (5 months). The 50 trial cross-validation was performed and the median, the lower and upper 95% confidence interval (LCI and UCI) of MAE values were determined (Table 5.1). Our proposed model outperforms the previous studies on gestational age estimation, which were based on 1D-DUS signals and maternal blood pressure and heart rate [248] and 1D-DUS signals with simultaneously recorded fetal electrocardiogram (ECG) [170, 171]. It should be noted that although using simultaneously recorded ECG or maternal blood pressure can improve the performance of the gestational age estimation [248, 201], for the application of interest, requiring an additional device for blood pressure or ECG recordings significantly complicates the use and raises the cost of the smartphone-mediated perinatal screening system.

Table 5.1: Mean absolute errors of the 50-trial five-fold cross validation for the CLSTM in months. Error is reported as lower, median and upper 95% confidence interval (LCI, median, UCI) for GAs of 5-9 months, together with the average over all months tested (All).

	Gestational Age (months since reported LMP)					
	5	6	7	8	9	All
Error	(1.9, 1.98, 2.1)	(0.7, 0.72, 0.8)	(0.4, 0.45, 0.5)	(0.4, 0.48, 0.4)	(0.9, 0.98, 1.1)	0.71

5.7 Discussion and conclusion

This work represents the first attempt to estimate gestational age from only Doppler signals, and outperforms previous attempts based on multiple signals (Doppler plus electrocardiogram [170] or Doppler plus blood pressure [248]). Since the error is close to the quantization of the labels, future improvements will require more accurate GA labels, collected using Doppler imaging in the first trimester.

Acknowledgements

GC and NK acknowledge the support of the National Institutes of Health, the Fogarty International Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number 1R21HD084114-01 (Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala). GC has financial interest in Alivecor Inc., and receives unrestricted funding from the company. GC is the CTO of Mindchild Medical Inc. and has ownership interests in Mindchild Medical Inc. RS also has equity interests in Mindchild Medical Inc. The research presented here is unrelated to, and is not funded by Mindchild Medical Inc. or Alivecor Inc.

Chapter 6

Detection of noisy gestational age recordings

6.1 Abstract

The growing application of mobile health (mhealth) technology for maternal and child health monitoring emphasizes examination of the contextual factors and interventional outcomes to improve the prototype design. Previously, we introduced an affordable fetal monitoring system aimed to assist low-literacy traditional birth attendants (TBAs) in rural Guatemalan communities. This mHealth fetal monitoring system consists of a low-cost 1D Doppler device connected to a smartphone running a mobile application. Before recording Doppler ultrasound signals, the app guides the user to enter the anticipated gestational age in months based on the last menstrual period. Further analysis has been done to evaluate fetal development based on fetal cardiac recordings and gestational age labels using machine learning and signal processing methods. In this study, we examined the reliability of the recorded gestational age labels based on the distribution of the times of recording and the values recorded in order to improve the estimates. Using multiple observation fusion we found that,

there is a correlation between gestational age and uncertainty in labels. While further analysis should be done to detect source of the error, we hypothesize that it might be due to recalling the conception date in last weeks of pregnancy. The proposed method will be used as a metric for detecting noisy labels for future analysis.

6.2 Introduction

Healthcare challenges in low and middle-income countries (LMICs) have been the focus of many digital initiatives that aim to improve both access to healthcare and quality of care. However, moving beyond the initial phase of piloting and experimenting needs effective scaling and integration to provide sustainable benefits to the healthcare system. In particular, to evaluate the effectiveness of an intervention in LMICs, it is essential to investigate sources of inaccurate or unreliable recordings and take the necessary steps to provide users with real-time feedback for immediate corrective action or modify the design of the monitoring system to avoid incorrect or noisy recordings. Unreliable data could be due to the user's habitual mistakes or equipment malfunctions, leading to incorrect insight and faulty predictions.

In this work, we focused on the fetal maternal health monitoring system provided for traditional birth attendants in rural Guatemala [166, 234, 237, 167, 251], where the perinatal morbidity and mortality rates are the highest in Latin America [277, 270]. Maternal and neonatal health is a key concern of the global healthcare community. Although advances in healthcare technology and delivery have reduced overall morbidity and mortality rates in many countries, neonatal and maternal outcomes have not improved proportionally. It is estimated that each year over 287000 women die of pregnancy-related causes [95, 277], and there are 2.6 million stillbirths and 2.8 million early neonatal deaths worldwide [38, 146, 262]. Limited funding for medical technology and poor infrastructure for delivering and maintaining technology

severely limit medical decision support in these areas. To address these challenges, a low-cost smartphone-based monitoring system was proposed, including peripheral sensors, such as a handheld Doppler for the identification of fetal compromise. This system is provided to assist traditional birth attendants who were trained to use the developed mobile app and the hand-held 1D Doppler device for recording cardiac activity. Before recording signals, they also entered the estimated gestational age into the app in months, based on the last menstrual period (LMP). The precision of the gestational age labels could be affected by different factors in the process of recording such as error in recalling the date of LMP or entering the correct month. This error will significantly affect further analysis of fetal health and development. Therefore, we proposed a method to detect reliable gestational age labels and improve the analysis of deep learning based gestational age estimation model [131].

In post processing steps that we perform on clinical data acquired in high income countries we can discard the poor quality records, record them again or switch to a more reliable monitor. However, when we propose solutions for LMICs with limited and overloaded medical resources, misreported values, low quality signals and images become an integral part of the problem. Therefore, it is necessary to provide users with real time feedback on data quality during the patient information acquisition. As part of the quality assessment process, we had previously presented the fetal Doppler signal quality evaluation method to detect the type of noise and classify the signal into interference, silent, talking, poor and good quality classes [250, 251]. Despite all the aforementioned challenges, it had been demonstrated that midwifes could feasibly collect quality one-dimensional fetal Doppler ultrasound data using a low-cost device and smartphone [122]. In this work, we looked into the quality assessment from a different angle and proposed a systematic approach to evaluate the accuracy of the gestational age labels recorded by the Guatemalan TBAs.

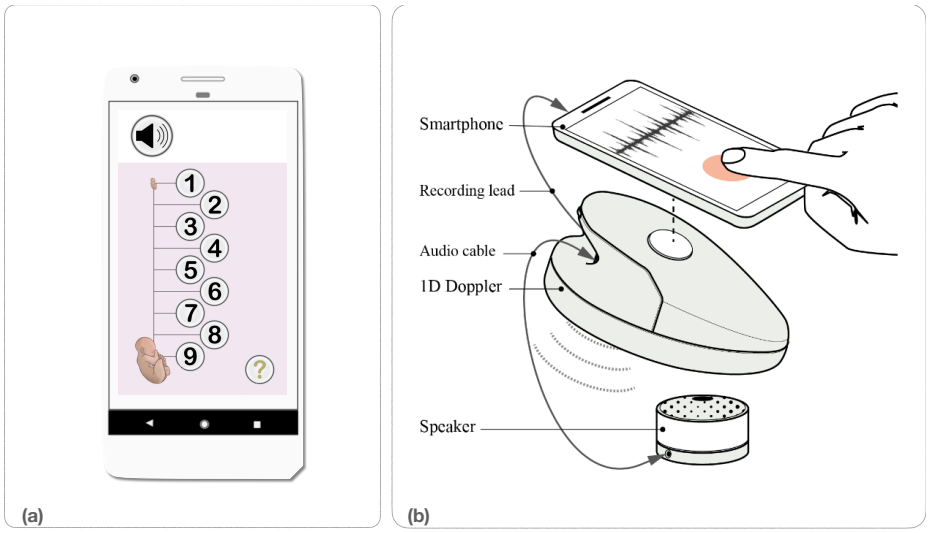


Figure 6.1: Data collection. a) Gestational age is recorded in months on the phone using the developed mobile app. b) The raw 1D ultrasound is captured using Doppler transducer (adapted from [251]).

6.3 Method

6.3.1 Data acquisition

The data was collected as part of a randomized control trial, conducted in rural highland Guatemala [167, 251]. The recorded data includes multimodal data from visits that were conducted inside the patient's home and contains recording of 2769 patients. The primary objective of the developed mHealth perinatal monitoring system is to enable recording of fetal cardiac activity onto a mobile phone for further processing and fetal risk analysis. In order to detect growth restriction and monitor fetal neural development, the gestational age at the time of visit was also recorded (figure 6.1). The 1D-DUS device was an AngelSounds Fetal 1D-DUS JPD-100s (Jumper Medical Co., Ltd., Shenzhen, China) with an ultrasound transmission frequency of 3.3 MHz. Data were captured using a bespoke Android client at 44.1 kHz, using a low-cost smartphone (Samsung S3 mini) and stored as uncompressed WAV files at 7056/s bits) [237].

6.3.2 Maternal healthcare assessment

Monitoring the enrollment and frequency of visits can help to detect underlying barriers to care and access for patients and is one of the important aspect of managing the health research since lengthy assessment intervals could miss essential health problems.

The distribution of the visits was determined to probe the frequency of visits in this study. The Poisson distribution was fitted to the histogram of the visits. In fitting a Poisson distribution to the data, each of the number of visits has the probability mass function:

$$\pi_k = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{6.1}$$

In order to fit the Poisson distribution, value for λ was estimated from the observed data which is also the expected value and variance of a Poisson distributed random variable.

6.3.3 Multiple measurement fusion

A pregnant woman may be admitted at the clinic or by the midwife several times throughout the pregnancy, resulting in multiple traces of the fetal Doppler recording at different gestational ages. Let $a(t)$ denote the reported gestational age acquired during a clinical visit on date t. Therefore, the anticipated conception date can be calculated by equation $c = t - a(t)$. Next suppose that a total number of N Doppler signal have been recorded from a pregnant woman at dates t_k $(k = 1, \ldots, N)$, resulting in the gestational age labels $\{a(t_1), \ldots, a(t_N)\}$. Each of these labels gives an estimate of the "true conception date" c, i.e.,

$$c_k = t_k - a(t_k), \quad k = 1, \ldots, N \tag{6.2}$$

The N conception date estimates can be merged together to obtain a more accurate estimate of the true conception date. For each patient a set of estimated conception dates is available which were subject to some kind of random error, either due to error in recalling the LMP date or entering incorrect gestational age label. In order to reduce the effect of outliers in the final estimation we took the median of the resulting measures and denoted by $\hat{c}$. After estimating the conception date using multiple measurement fusion, the distance from the estimated date has been used as a metric to score the reliability of the recorded gestational age.

6.4 Results

6.4.1 Distribution of the visits

The recording of the data has been started since 2016. In this section we investigated the number of visits for patients in the study at each year. In figure 6.2 we illustrated the histogram and fitted Poisson distribution. This emphasizes the fact that number of visits follow the Poisson distribution and the resultant long-tail distribution is based on a nature of this count data. Therefore, large number of visits should not be considered as an outlier in analysis of gestational age label.

6.4.2 Evaluation of gestational age labels

In order to use multiple data fusion to evaluate the accuracy of gestational age labels preprocessing of data was needed. Hence, the data was filtered based on three conditions which are prerequisite to the analysis of quality of the labels. First, the existence of gestational age labels, second the existence of date of visit and the last one, existence of more than one visit after filtering based on first two conditions. The filtered data includes 565 patients which had all the information needed for evaluating the labels. In figure 6.3 distribution of visits and labels are provided after filtering.

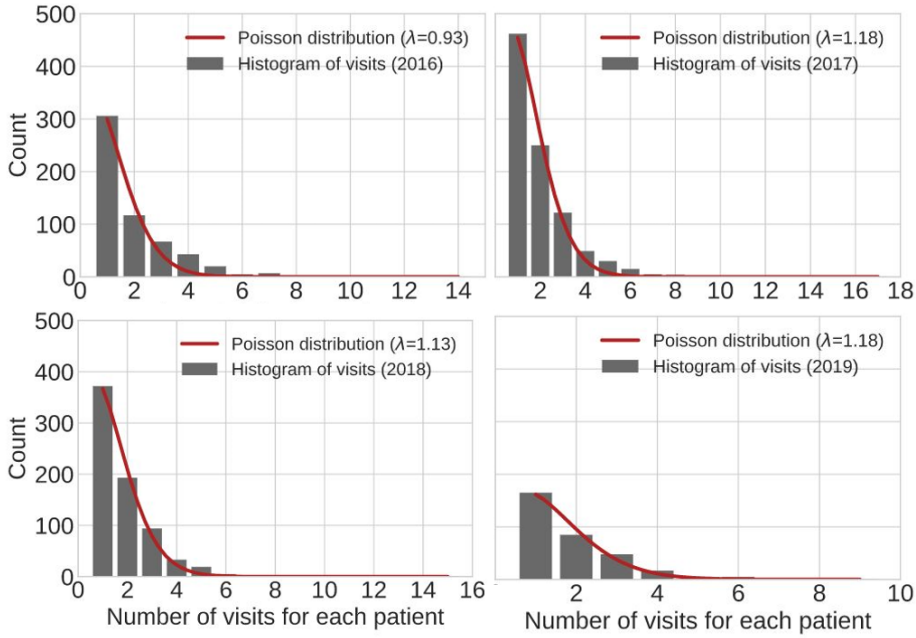


Figure 6.2: Distribution of number of visits from 2016 to 2019.

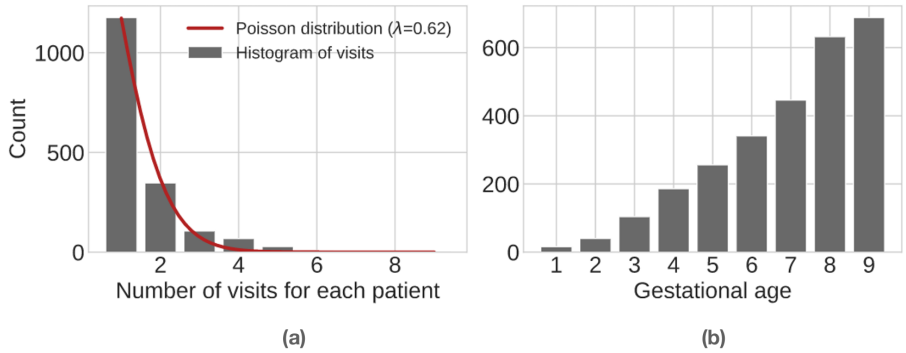


Figure 6.3: Distribution of the data for analysing the quality of the gestational age labels. a),b) illustrate the histogram of visits and gestational age labels after filtering the data.

To analyze the reporting gestational age we have evaluated the distribution of the error based on variables of the study as follows:

- Patient and midwife IDs.
- Date and month of visit.
- Label of gestational age.

6.5 Discussion and conclusion

The recorded data in visits of the study may not include all the expected modalities such as Doppler recording and gestational age label. In the processing of Doppler signal to estimate gestational age we will consider the error in recorded gestational age labels based on the proposed evaluation metric. The logic behind this process is to determine which inputs to use as a clean data to train and evaluate the deep learning model. It should be noted that we do not have enough information to evaluate the accuracy of the gestational age label for patients with just one visit. Therefore, it is not possible to draw a concrete conclusion based on this part of the data.

Acknowledgements

GC and NK acknowledge the support of the National Institutes of Health, the Fogarty International Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number 1R21HD084114-01 (Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala). GC has financial interest in Alivecor Inc., and receives unrestricted funding from the company. GC is the CTO of Mindchild Medical Inc. and has ownership interests in Mindchild Medical Inc. RS also has equity interests in Mindchild Medical Inc. The research presented here is unrelated to, and is not funded by Mindchild Medical Inc. or Alivecor Inc.

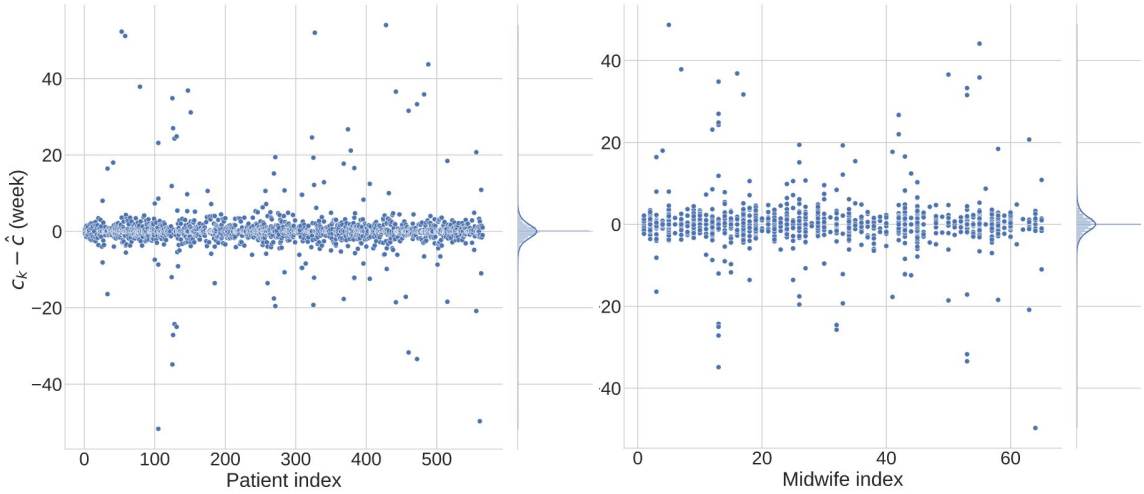


Figure 6.4: Difference of estimated conception date ($\hat{c}$) and dates from multiple observations (c_k) based on patient index and midwife index.

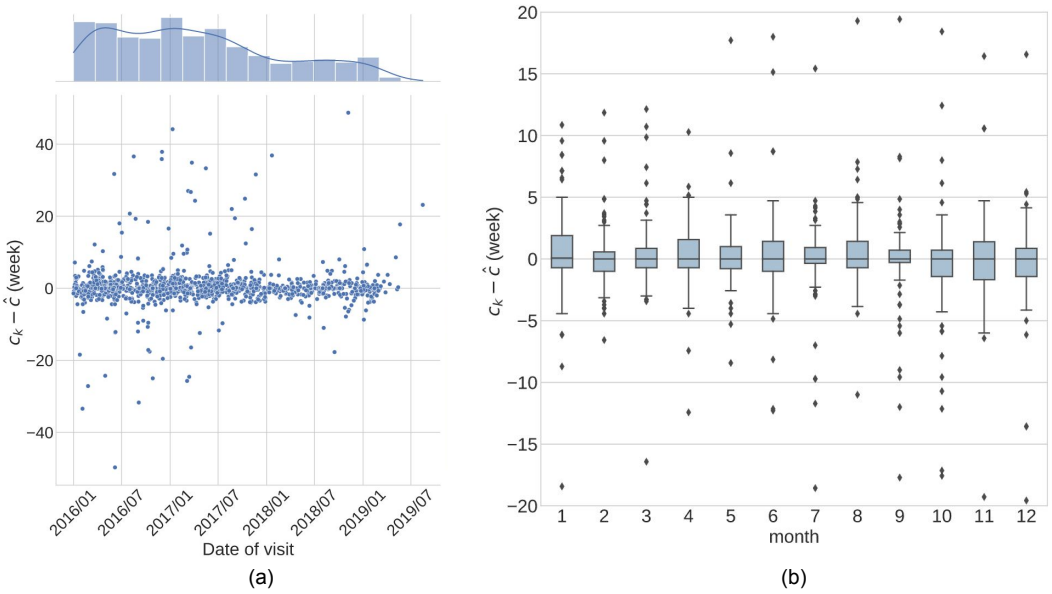


Figure 6.5: Difference of estimated conception date ($\hat{c}$) and dates from multiple observations (c_k) a) based on date of visit and b) based on month of visit.

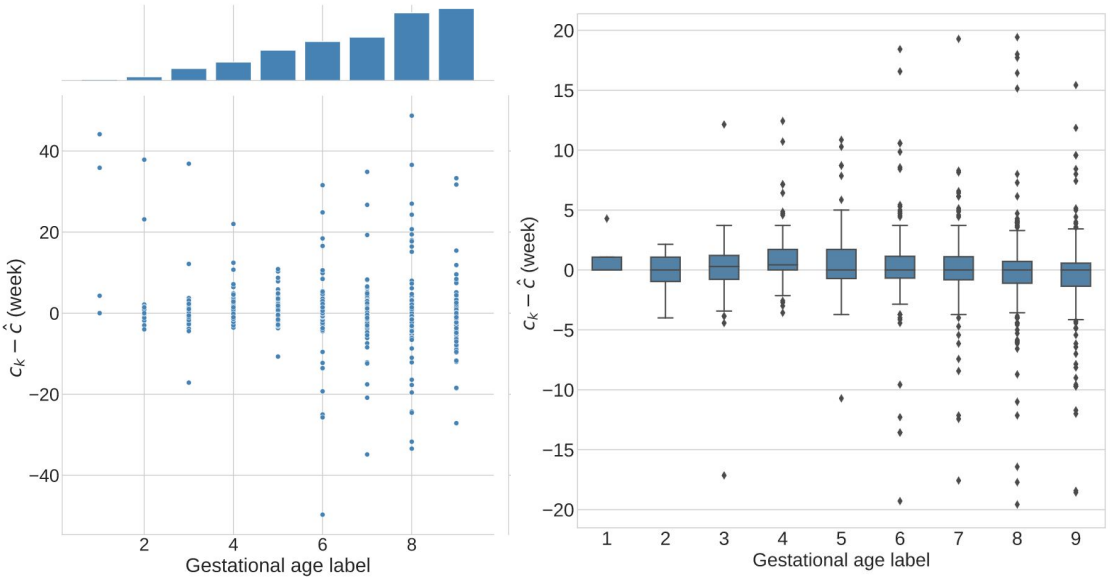


Figure 6.6: Difference of estimated conception date ($\hat{c}$) and dates from multiple observations (c_k) a) based on date of visit and b) based on month of visit.

Chapter 7

Hierarchical attention network for gestational age estimation

7.1 Abstract

Assessing fetal development is essential to provide a standard of care medical management for both mother and fetus. In low- and middle-income countries, conditions that increase the risk of intrauterine growth restriction (IUGR) are often more prevalent. In these regions, barriers to access health care and social services exacerbate the issue. This work introduces a series of machine learning algorithms for use on an affordable hand-held Doppler and smartphone-based system for estimating gestational age (GA), and by inference, IUGR.

Doppler ultrasound signals were collected from 401 pregnancies between 5 and 9 months GA by traditional birth attendants in highland Guatemala. A subset of the data were labelled for quality and a method for automatically classifying quality was implemented.

A model for GA estimation was then proposed based on a hierarchical sequence learning method with attention. This resulted in an state-of-the-art performance,

with an average GA estimation error of 0.72 months. This is close to the theoretical minimum for a quantization level of one month. Using a balanced batch generator with assigned weights based on the inverse class frequency, the error in estimating label five was reduced from 1.77 to 1.21 months.

The methods developed in this thesis are readily deployable to provide AI-guided diagnostics This AI-enabled edge computing device is intuitive to use, even for low-literacy populations.

7.2 Introduction

Accurate estimation of gestational age (GA) is crucial to identify infants at risk for adverse health outcomes. GA is a proxy for fetal development and is used for delivery scheduling, and preterm birth management [5]. The prevalence of growth restriction is known to be high where an estimated twenty million infants are born globally with low birth weight (less than 2500 g) every year, and the majority are born in low-and middle-income countries (LMICs) [264]. Fetuses with intrauterine growth restriction (IUGR) conditions are more vulnerable to mortality and morbidity in the neonatal period and beyond. Therefore, early prediction of IUGR could help manage the condition and lower mortality risk.

Maternal and child mortality is an important public health issue since most of them are preventable by addressing inequalities that affect maternity care. Almost half of women in low-and middle-income countries (LMICs) do not receive adequate antenatal care, and worldwide an estimated two million early neonatal deaths occur annually in these areas primarily due to lack of access to quality care [149, 77]. In high-income countries ultrasound imaging is currently most frequently used for fetal health monitoring and estimating GA. Nonetheless, the cost of purchase, the technical skills required for maintenance and the user-dependent accuracy have limited the

application of this technique in resource-limited settings [189]. Therefore, low-cost alternative methods are used in LMICs to estimate GA. A common method used for GA estimation is the last menstrual period (LMP), in which a 28-days menstrual cycle is assumed. Although previous studies have criticized LMP due to the inconsistency in the menstrual cycle length [68], and the difficulty to recall the day of the last menstrual period [16], the LMP method has shown to be a somewhat useful method for LMICs, particularly in rural areas lacking medical equipment. However, recently the increasing proliferation of mobile technology is bringing up new opportunities to permit safe, accessible, cost-effective, and coordinated maternal health care, which could reduce many of the pregnancy complications that afflict populations in LMICs.

Fetal cardiac assessment is crucial to identify high risk fetuses [154]. The Autonomic Nervous System (ANS) regulates fetal heart rate [218, 261], which modifies FHR dynamics during pregnancy. Therefore, FHR is associated with fetal development and gestational age, which could facilitate the detection of pathological fetal development [258]. Studies on the detection of growth restriction using FHRV showed that IUGR fetuses have a lower percentage of heart rate variability compared with the normal population [238]. It should be noted that FHRV is influenced by not only maturation but also the fetal behavioral state. Originally, the behavioral states are defined as quiet sleep, active sleep and active awakeness after 32 weeks of gestation. Before then, it is only possible to distinguish between quiet and active states [214]. Besides FHRV, another indicator of fetal state can be seen as a concept based on movement patterns and accelerations. In conclusion, FHRV markers, together with information on behavioral states, can contribute to the detection of fetal development issues. Cardiotocography is an inexpensive Doppler-based method that is routinely performed during pregnancy for fetal heart monitoring. However, it has low specificity and is endowed with an auto-correlation of the beats, which reduces the resolution of the estimated heart rate. Another type of device for recording fetal cardiac activity

using Doppler technology is in-home fetal Doppler transducers. This non-invasive and low-cost technique can be easily adapted to connect to mobile devices such as smartphones for recording and processing, motivating their use in mobile-health (mhealth) systems for risk screening in low-resource environments [237].

Using the Doppler technique, the flow of blood through the heart's chambers and valves can be captured. Therefore, analyzing 1D-DUS in the time and frequency domain provides comprehensive information regarding fetal health. However, the morphology of 1D-DUS signals is highly variable. This intra- and inter-subject variability is caused by fetal movement and location of the transducer. Therefore, learning heart rate patterns from 1D-DUS is challenging due to transient nature and changes in the statistical characteristics of the signal. Machine learning approaches, especially deep learning techniques, have achieved significant progress in processing physiological signals. This work presents the Hierarchical deep sequence learning model to estimate gestational age from fetal 1D-DUS recordings. The network consists of two levels of attention mechanism to capture long and short-term variability of cardiac activity. The hierarchical attention network has been introduced for document classification [283] and designed to capture two basic insights about document structure. First, a hierarchical structure of the documents, a document representation is constructed by first building the representations of sentences and then aggregating those into a document representation. Second, the fact that different words and sentences in a document are deferentially informative.

7.3 Related Work

Previous works on GA estimation are based on finding a relation between fetal development and FHRV metrics. Specifically, it has been shown that using FHRV parameters extracted from magnetocardiographic recordings as an input of regres-

sion model fetal maturation age can be assessed [110, 240]. However, this approach requires high resolution fetal magnetocardiographic recordings which is costly and non-portable equipment, making its use in LMICs impractical.

In earlier works, Marzbanrad *et al.* presented a method on estimating GA using a step-wise regression based on cardiac wall intervals derived from 1D-DUS and electrocardiagram signals [170]. In further work, Marzbanrad *et al.* improved the estimation accuracy by incorporating 1D-DUS and fECG quality assessment algorithms and excluding poor quality signals [171]. Recently, Valderrama *et al.* presented a study on using fHRV indexes derived from 1D-DUS and maternal blood pressure to estimate GA using support vector regression [248]. Although the presented Doppler based methods achieved significant results, they need additional recordings such as electrocardiagram signals or maternal blood pressure, which increases costs and complicates implementation, particularly in LMICs.

Deep learning models with the capability of automatic feature extraction provide a significant performance in processing of physiological data. Most of the presented algorithms on cardiac signal classification using deep learning approach in the literature are based on ECG signals. Recent works on deep sequence learning with attention based models improved the interpretability and performance of the learning process. To provide an interpretable model with high performance for automatic estimation of gestational age, in this study, we proposed a deep learning model powered by hierarchical attention networks.

7.4 Method

In this paper, we focus on modeling of the fetal cardiac sequence to estimate gestational age. Let $x(t; d)$ denote the time series of a 1D-DUS signal with discrete time index t, acquired during a clinical visit of a pregnant woman on date d. For simplicity,

the date and gestational age are represented in units of weeks. The "true gestational age" at date d is denoted $a(d) = d - c$, where c is the *date of conception*, while the presumed or reported gestational age is $\tilde{a}(t) = t - \tilde{c}$, where $\tilde{c}$ is the *anticipated conception date*. Therefore, the presumed (anticipated) and true gestational ages can be related as follows:

$$\tilde{a}(t) = a(t) + \eta \tag{7.1}$$

where $\eta = c - \tilde{c}$ is the gestational age presumption error, which without additional priors (such as 2D-Doppler) remains an unknown stochastic constant over pregnancy. The error η accounts for lack of knowledge of the last menstrual period and uncertainties in the exact ovulation, intercourse and conception dates. We further denote the 2-dimensional scalogram feature extracted from the 1D-DUS by $\mathbf{f}(x(t;d)) \in \mathbb{R}^2$. The objective is to design a deep network that estimated the true gestational age from a single or a set of 1D-DUS acquired during pregnancy, i.e.,

$$\hat{a}(t) = \Gamma\left[\tilde{a}(t), \{\mathbf{f}(x(t;d_k))\}_{k=1}^{L}\right] \tag{7.2}$$

where $\hat{a}(t)$ is an estimate of the true gestational age, d_k $(k = 1, \ldots, L)$ denote the L dates that 1D-DUS is acquired from the pregnant woman, and $\Gamma(\cdot)$ denotes the 1D-DUS to gestational age transform that is learned by the neural network, as shown in Fig. 7.1. In this scheme, the presumed gestational age $\tilde{a}(t)$ is used for model training.

7.4.1 Hierarchical attention network for modeling long- and short-term temporal patterns

We leveraged hierarchical attention network to test the hypothesis that better representations can be obtained by incorporating knowledge of long-term and short-term fetal cardiac activity in the model architecture. The intuition underlying our model is that not all parts of a signal are equally relevant for estimating gestational age.

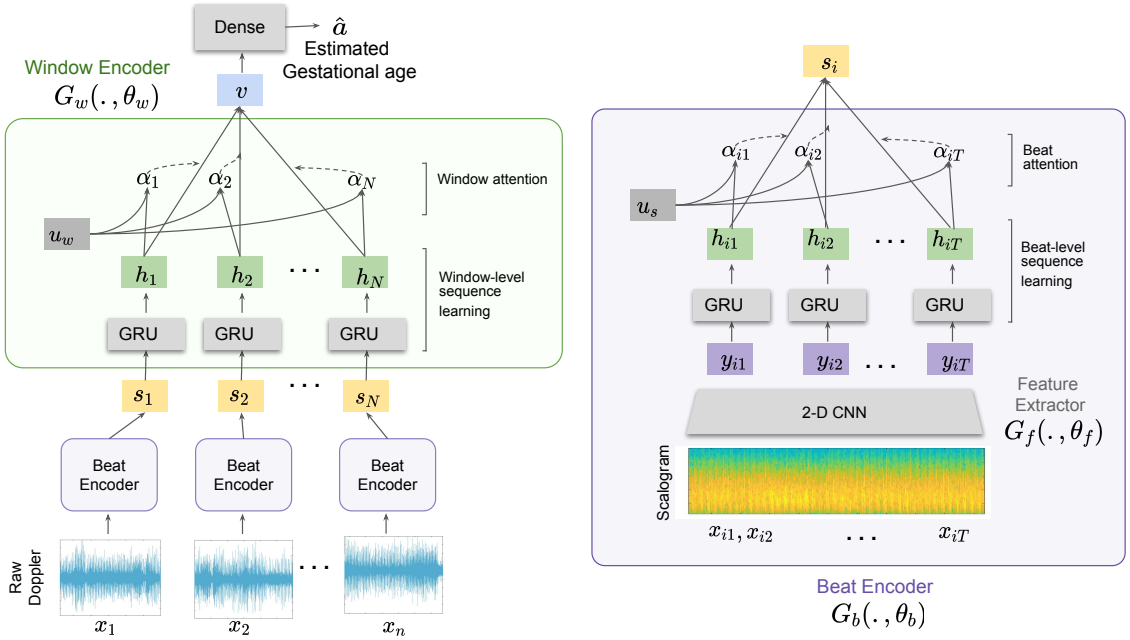


Figure 7.1: The architecture of the proposed hierarchical attention network. It contains three main components: 1) convolutional feature extractor, 2) beat encoder and 3) window encoder.

Determining the relevant sections involves modeling the sequence of windows which are the group of the beats. This model includes two levels of attention mechanisms, one at the time sample level focusing on the scalogram of the Doppler signal and one at the window level. By applying successive attention layers the model attempts to learn which time samples and windows play the most important role in the signal classification. By determining the importance of segments of Doppler signals we can also attain better understanding about fetal movements and behavioral states.

The network $\Gamma(\cdot)$ has three components, feature extractor $(G_f(.,\theta_f))$, beat encoder $(G_b(.,\theta_b))$ and window encoder$(G_w(.,\theta_w))$. Both beat and window encoders include attention layer. The feature extractor is a space invariant neural network that learns a representation based on training data by finding a robust transformation. The beat encoder network is a recurrent network that learns the dynamic of set of beats. Finally, the window encoder learns the relation of segments of multiple beats.

7.4.2 Sequence Encoder

In order to model the sequence of beats and segments, gated recurrent units (GRU) [24] are used. The GRU uses a gating mechanism to track the state of sequences without using separate memory cells. We denote input vector at time t as x_t, one can adapt the GRU architecture as:

$$\begin{aligned} z_t &= \sigma(U_z x_t + W_z h_{t-1} + b_z), \\ r_t &= \sigma(U_r x_t + W_r h_{t-1} + b_r), \\ \tilde{h}_t &= \phi(U_h x_t + W_h(r_t \odot h_{t-1}) + b_h), \\ h_t &= z_t \odot \tilde{h_t} + (1 - z_t) \odot h_{t-1}, \end{aligned} \tag{7.3}$$

where r_t is a reset gate and z_t update gate. r_t decides how much information should be preserved and z_t decides the contribution proportion of the past and new information. σ and ϕ are point-wise nonliniarity, $\odot$ is point-wise product and W, U, b are parameters of the model. Both beat encoder $G_b(., \theta_b)$ and window encoder $G_w(., \theta_b)$ networks include GRU layers.

7.4.3 Hirarchical Attention

As mentioned earlier, the hierarchy in the network tries to incorporate long and short-term dynamics in 1D-DUS. It is obvious that some parts of the signal are more involved in a given task due to fetal behavioral states, movement patterns and quality of the signal. Therefore, this model utilizes two levels of attention mechanism along with hierarchical training of beat-level and window level networks.

Suppose that h_{it} is a hidden representation of the time sample t in window i in vector space, the attention layer in $G_b(., \theta_b)$ network first projects h_{it} into hyperbolic space (u_{it}). Then, it combines the components of u_{it} according to their relevance to the problem and estimate the normalized importance weight α_{it} through a softmax

function. After that, the weighted sum of the time sample representations creates the window vector s_i.

$$\begin{aligned} u_{it} &= \tanh(W_b h_{it} + c_b), \\ \alpha_{it} &= \frac{\exp(u_{it}^T u_b)}{\sum_\tau \exp(u_{i\tau}^T u_b)}, \\ s_i &= \sum_t \alpha_{it} h_{it}. \end{aligned} \tag{7.4}$$

The Window vectors $s_1, ..., s_i$ are then fed to the $G_w(., \theta_w)$ network. The window-level attention gets hidden representation of windows after processing in GRU layer. In Eq 7.5, v is a high level representation and summarizes the information in one recording of 1D-DUS. The window-level attention mechanism works as follow:

$$\begin{aligned} u_i &= \tanh(W_w h_i + c_w), \\ \alpha_i &= \frac{\exp(u_i^T u_w)}{\sum_i \exp(u_i^T u_w)}, \\ v &= \Sigma_t \alpha_i h_i. \end{aligned} \tag{7.5}$$

7.5 Experimental set-up

7.5.1 Data

1D-DUS recordings were captured from 401 pregnant women (415 visits and 693 recordings) at 5 to 9 months of gestation. The data includes 15, 77, 162, 186, 253 recordings corresponding to gestational ages of 5, 6, 7, 8 and 9 months, respectively. The data were collected as part of a randomized control trial, conducted in rural highland Guatemala [167, 251]. The dataset includes 1D-DUS signals recorded by traditional birth attendants, who were trained to use the hand-held 1D-DUS device and were provided with a mobile application. The app guides the user to enter the anticipated gestational age in months based on the last menstrual period before

recording the 1D-DUS signals. The 1D-DUS device is an AngelSounds Fetal 1D-DUS JPD-100s (Jumper Medical Co., Ltd., Shenzhen, China) with an ultrasound transmission frequency of 3.3 MHz. Data were captured using a bespoke Android client at 44.1 kHz, using a low-cost smartphone (Samsung S3 mini) and stored as uncompressed WAV files at 7056/s bits) [237]. The first five minutes of each recording are used in this study.

7.5.2 Preprocessing

Given the nature of the physiological time-series, 1D-DUS signals are corrupted with internal and external interference such as respiration, movement, and environmental noise. In this work, a second-order band-pass Butterworth filter is used to reduce the noise. By observing the frequency components of the 1D-DUS signals, the cut-off frequencies are set to 25 and 600 Hz, corresponding to cardiac oscillations. After the preprocessing steps, scalogram of Each 20-second window is generated followed by logarithmic scaling.

7.5.3 Network implementation

Feature extractor network ($G_f(., \theta_f)$) gets the scalogram of the signal and consists of three layers of 2-D convolutional neural network (kernel size=(3,3)). Each layers is followed by batch normalization, rectified linear (ReLU) units, and max pooling units. A mean absolute error (MAE) is used as a loss function and mini batch stochastic gradient decent (SGD) is leveraged to optimaze the parameters of the network. To address the class imbalance, balanced batch generator method is employed to improve the generalization on less frequent categories. In addition, weights are assigned to training samples proportional to the inverse class frequency. The proposed method is implemented in tensorflow 2.0 and Python3.

Table 7.1: Mean and standard deviation of the MAE (Mean (STD)) for gestational ages of 5-9 months, together with the average over all months tested (All).

	Gestational Age (months since reported LMP)					
	5	6	7	8	9	All
$\text{Error}_{\text{BBG}^*}$	1.77 (0.80)	0.98 (0.62)	0.65 (0.41)	0.40 (0.41)	0.87 (0.51)	0.72
$\text{Error}_{\text{BBG-SW}^{**}}$	1.21 (0.81)	0.93 (0.64)	0.69 (0.43)	0.43 (0.46)	0.94 (0.54)	0.74

* BBG: Balanced batch generator $\geq$ 24 weeks.
** BBG-SW: Balanced batch generator and assigning sample weight.

7.6 Evaluation metrics

To evaluate the performance of the model, MAE of estimating GA based on reported LMP are determined. In addition, box plots of errors are provided to show the range of error and outliers for each label. Results of using balanced batch generator and assigning sample weights are provided in section 7.7. In addition, to assess the performance of the model based on quality of labels, we have used the approach presented in 6 and created a set of clean labels to train and test the model. The cross validated results of estimating gestational age on clean labels using the attention based model and the model presented in chapter 5 are also provided in the result section.

7.7 Results

Stratified five-fold cross-validation is used to assess the performance of gestational age estimation. Training strategies for imbalanced dataset and effect of label noise are considered in this work. Table 7.1 shows the MAE of estimating GA in each month of pregnancy based on reported LMP. The results of the experiments show that, assigning sample weights in the training process reduced MAE of estimating label five from 1.77 to 1.21 and increased the error in estimating label nine from 0.87 to 0.94. Figure 7.2 shows the result of gestational age estimation using the defined training strategies for imbalanced data.

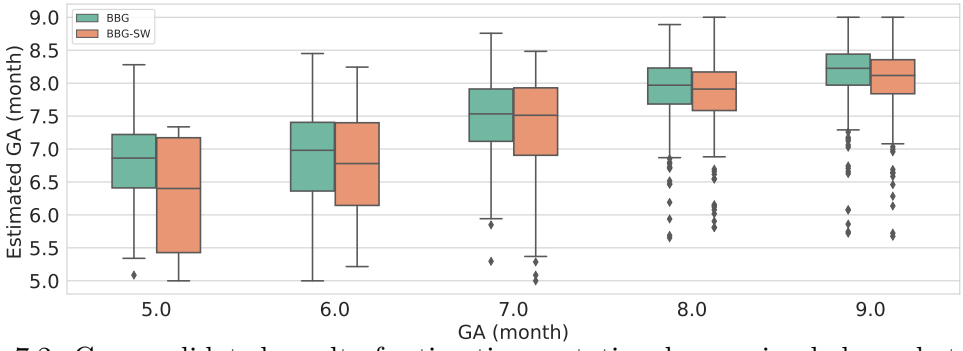


Figure 7.2: Cross-validated result of estimating gestational age using balance batch generator (BBG) and balance batch generation following by assigning sample weight (BBG-SW)

Table 7.2: Mean and standard deviation of the MAE (Mean (STD)) for gestational ages of 5-9 months, together with the average over all months tested (All).

	Gestational Age (months since reported LMP)					
	5	6	7	8	9	All
$\text{Error}_{\text{CLSTM}^*}$	1.83 (0.22)	1.21 (0.57)	0.60 (0.40)	0.66 (0.48)	0.84 (0.58)	0.80
$\text{Error}_{\text{HAN}^{**}}$	1.29 (0.46)	0.81 (0.73)	0.7 (0.57)	0.36 (0.32)	0.68 (0.57)	0.64

* CLSTM: Convolutional LSTM.
** HAN: Hierarchical Attention Network.

In another experiment, we tested the performance of the model on a set of data with clean labels. The data includes 5, 26, 56, 49 and 107 recordings corresponding to gestational ages of 5, 6, 7, 8 and 9 months, respectively. The results of the hierarchical attention network and convolutional LSTM model (chapter 5) are provided in this section. As it has been shown in table 7.2 and figure 7.3 the proposed hierarchical attention network outperforms the previously presented model based on convolutional LSTM network.

7.8 Discussion and conclusion

In this work, a hierarchical sequence learning with attention mechanism was proposed for the fetal gestational age estimation. The quality of the fetal cardiac signals and behavioral state of fetus at the time of recording of the data are essential factors which should be considered for estimating gestational age. The proposed model, weights the

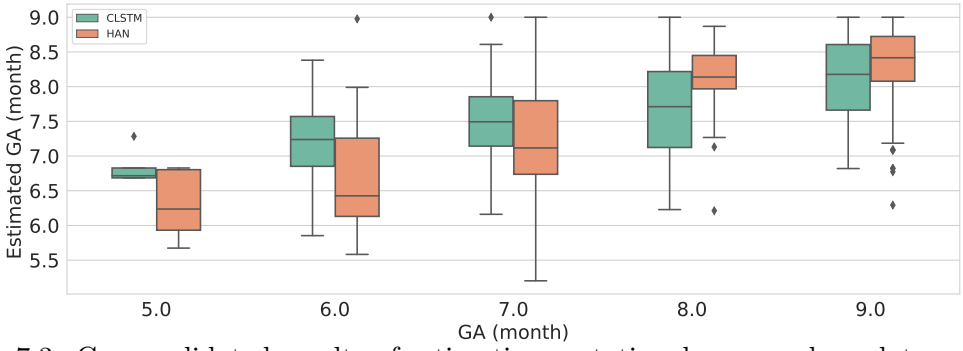


Figure 7.3: Cross-validated results of estimating gestational age on clean data using hierarchical attention network (HAN) and convolutional LSTM (CLSTM) networks.

important segments and time samples of the data according to the defined task using two levels of attention layers. In addition, since the imbalanced dataset used in this work could affect the generalizability of the model on less frequent labels, imbalanced learning strategies such as balanced batch generator and assigning sample weights were employed. The proposed method achieves the state-of-the-art performance on estimating gestational age using only Doppler signals.

Acknowledgements

GC and NK acknowledge the support of the National Institutes of Health, the Fogarty International Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number 1R21HD084114-01 (Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala). GC has financial interest in Alivecor Inc., and receives unrestricted funding from the company. GC is the CTO of Mindchild Medical Inc. and has ownership interests in Mindchild Medical Inc. RS also has equity interests in Mindchild Medical Inc. The research presented here is unrelated to, and is not funded by Mindchild Medical Inc. or Alivecor Inc.

Chapter 8

Discussion and conclusions

8.1 Summary and contributions

The work presented in this thesis addressed assessing fetal health using a low-cost monitoring system appropriate for low- and middle-income countries (LMICs) to enhance healthcare for disadvantaged populations by developing AI-enabled edge computing devices. Specifically, the monitoring system collects 1D-Doppler ultrasound signals (1D-DUS) of fetuses using a mobile health system previously introduced by Stroux *et al.* [234, 238] in a Guatemalan highland rural community. Theoretically, this thesis focuses on extracting temporal and spectral features from 1D-DUS signals to detect pregnancy complications using signal processing and machine learning methods. Tracking gestational development is crucial since growth restriction conditions increase the risk of perinatal mortality. In rural Guatemala, the rate of fetal maternal mortality is alarmingly high [244] and a leading contributor to this burden is IUGR [151], which is a slowdown of fetal growth leading to low birth weight. Low birthweight, in turn, is associated with lower neurodevelopmental scores and other adverse sequelae [182, 153]. Automating the detection of critical cases of restricted-growth fetuses can assist in managing timely referrals and interventions. One of the

approaches for assessing fetal development is fetal heart rate (FHR) monitoring which is widely used and recommended by obstetrical societies [154].

This work aimed to underline the importance of taking challenges in LMICs into account for designing a prototype, collecting data, and developing machine learning methods. Chapter 4 proposed an unsupervised probabilistic segmentation method to estimate fetal heart rate variability metrics from 1D-DUS signals. This method leveraged the embedding method to reduce the effect of highly variable patterns in 1D-DUS signals. Then the transformed data was fed to the hidden semi-Markov model segmentation model to detect beat onsets. Previous methods of FHR estimation using 1D DUS focused on two alternative approaches with associated weaknesses. The first approach is supervised and requires labeled data to train a segmentation model [173, 172] and is therefore dependent on the quality of the labels and the similarity between the labeled data and the new data encountered. The second approach ignores beat events and involves FHR estimation using fixed temporal windows of DUS data. The most popular approach in this latter category is based on an auto-correlation method [252] which provides a lower resolution for heart rate timing estimation than segmentation approaches due to averaging over a few seconds. Chapter 5 proposes an end-to-end learning approach to extract features from 1D-DUS signals and estimate gestational age. The recurrent network was used for deep sequence learning and extracting time dependency in cardiac signals. Previous works on assessing fetal development using 1D-DUS have leveraged additional recordings such as maternal blood pressure or fetal electrocardiogram, complicating the use and raising the cost of the smartphone-mediated perinatal screening system. This research presents the first attempt to assess fetal development using only Doppler signals. In chapter 6 we looked at the whole system from a different angle and proposed a systematic approach to evaluate the accuracy and reliability of the gestational age labels recorded in the Guatemala site. The results from this study were used in the gestational age

estimation model to improve the analysis. Finally, Chapter 7 introduced a novel deep hierarchical sequence learning model to incorporate short- and long-term heart rate patterns in 1D-DUS signals in estimating gestational age. The proposed model also consists of beat-level and window-level attention layers to emphasize the importance of different parts of the signal in estimating gestational age.

8.2 Limitations

The work presented in this thesis has some limitations. The methods presented in chapters 5, 6 and 7 were developed to track fetal gestational development and trained using the gestational age labels based on a last menstrual period. However, the last menstrual period is not totally accurate, thereby introducing noise at the labeling stage. In order to improve the accuracy and validate the labels, additional devices such as Doppler imaging is needed. However, this device is not available in the rural areas of Guatemala. Nevertheless, the approach presented in Chapters 6 and 7 can be easily adjusted to a dataset with more accurate gestational age labels such as those provided by Doppler imaging.

The deep learning approaches presented in this work are trained by backpropagation in a batch learning setting, which requires the entire training data to be made available prior to the learning task. Therefore, the developed methods were not tested online. The offline approach affects the scalability of the model for real-world scenario where new data arrives sequentially in a stream form.

8.3 Future work

The resultant system could empower women across LMICs, saving millions of lives a year. We therefore envision this system as a method of financially empowering the local communities and aim to develop a TPU-based version of the fetal health assess-

ment algorithms (using TensforFlow Lite). The TPU is an AI accelerator application-specific integrated circuit developed specifically for neural network machine learning, and is particularly suited to the TensorFlow software. The TPU will be integrated with the Doppler device and use the mobile phone as the receiving interface.

The proposed research is not limited to assessing fetal development, and could be extended to the detection of a wide range of maternal and fetal abnormalities using a multi-task learning scenario. In future works, we will develop methods to learn transferable and discriminative representations to discover the task relationships and jointly learn transferable features using deep networks.

Additionally, we will improve our methods by considering learning strategies in presence of label noise. Several methods have been proposed to make models more robust to noise such as noise transition matrix [194], robust losses [161], sample weighting [263], sample selection [185], meta-learning [203], we will also consider combined approaches to improve the performance of the model.

In summary, the next step is to address the label noise in optimizing the parameters of the model and expanding the proposed approaches to comprehensive fetal maternal health monitoring system for detecting wide range of pregnancy complications such as fetal growth restriction and hypertension in pregnancy.

Bibliography

[1] A. Z. Abuhamad and R. Chaoui. *A practical guide to fetal echocardiography: normal and abnormal hearts*. Lippincott Williams & Wilkins, 2012.

[2] P. C. Adithya, R. Sankar, W. A. Moreno, and S. Hart. Trends in fetal monitoring through phonocardiography: Challenges and future directions. *Biomedical Signal Processing and Control*, 33:289–305, 2017.

[3] H. M. Al-Angari, Y. Kimura, L. J. Hadjileontiadis, and A. H. Khandoker. A hybrid EMD-Kurtosis method for estimating fetal heart rate from continuous Doppler signals. *Frontiers in Physiology*, 8:641, 2017.

[4] A. Al-Zaben and A. Al-Smadi. Extraction of foetal ECG by combination of singular value decomposition and neuro-fuzzy inference system. *Physics in Medicine and Biology*, 51(1):137–143, 2006.

[5] G. R. Alexander, M. E. Tompkins, D. J. Petersen, T. C. Hulsey, and J. Mor. Discordance between LMP-based and clinically estimated gestational age: implications for research, programs, and policy. *Public Health Reports*, 110(4):395, 1995.

[6] Z. Alfirevic and J. P. Neilson. Doppler ultrasonography in high-risk pregnancies: systematic review with meta-analysis. *American Journal of Obstetrics and Gynecology*, 172(5):1379–1387, 1995.

[7] Z. Alfirevic, T. Stampalija, and T. Dowswell. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database of Systematic Reviews*, (6), 2017.

[8] L. D. Allan. *Manual of fetal echocardiography*. MTP Press, Lancaster, UK, 1986.

[9] L. D. Allan. A practical approach to fetal heart scanning. *Seminars in Perinatology*, 24(5):324–330, 2000.

[10] E. R. Allanson, M. Muller, and R. C. Pattinson. Causes of perinatal mortality and associated maternal complications in a South African province: challenges in predicting poor outcomes. *BMC Pregnancy and Childbirth*, 15(1):37, 2015.

[11] D. S. Alves, V. C. Times, É. M. A. da Silva, P. S. A. Melo, and M. de Araújo Novaes. Advances in obstetric telemonitoring: a systematic review. *International Journal of Medical Informatics*, 134(104004), 2019.

[12] I. Amer-Wåhlin, C. Hellsten, H. Norén, H. Hagberg, A. Herbst, I. Kjellmer, H. Lilja, C. Lindoff, M. Månsson, L. Mårtensson, and P. Olofsson. Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. *The Lancet*, 358(9281):534–538, 2001.

[13] American College of Obstetricians. ACOG practice bulletin. Antepartum fetal surveillance. Number 9, October 1999. Clinical management guidelines for obstetrician-gynecologists. *International Journal of Gynaecology and Obstetrics*, 68:175–185, 2000.

[14] M. Amin, M. Mamun, F. Hashim, and H. Husain. Separation of fetal electrocardiography (ECG) from composite ECG using adaptive linear neural network for

fetal monitoring. *International Journal of Physical Sciences*, 6(24):5871–5876, 2011.

[15] P. Anastasiadis, A. Kotini, P. Anninos, A. Adamopoulos, J. Sigalas, and N. Koutlaki. Chaotic and periodic analysis of fetal magnetocardiogram recordings in growth restriction. *Prenatal diagnosis*, 23(5):405–409, 2003.

[16] H. F. Andersen, T. R. Johnson Jr, J. D. Flora Jr, and M. L. Barclay. Gestational age assessment: II. Prediction from combined clinical observations. *American Journal of Obstetrics and Gynecology*, 140(7):770–774, 1981.

[17] F. Andreotti. *Extraction and detection of fetal electrocardiograms from abdominal recordings*. PhD thesis, Technische Universität Dresden, Leipzig, Germany, 2016.

[18] F. Andreotti, J. Behar, S. Zaunseder, J. Oster, and G. D. Clifford. An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms. *Physiological Measurement*, 37(5):627, 2016.

[19] N. Archer and N. Manning. *Fetal Cardiology*. Oxford University Press, Oxford, UK, 2009.

[20] K. Assaleh. Extraction of fetal electrocardiogram using adaptive neuro-fuzzy inference systems. *IEEE Transactions on Biomedical Engineering*, 54(1):59–68, 2007.

[21] A. M. Awiti, B. M. Shifferaw, M. B. Byamukama, R. Kizito, and C. Mwikirize. Design and implementation of an Android based digital fetoscope. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 152–155. IEEE, 2016.

[22] D. Ayres-de Campos and J. Bernardes. Twenty-five years after the FIGO guidelines for the use of fetal monitoring: Time for a simplified approach? *International Journal of Gynecology and Obstetrics*, 110(1):1–6, 2010.

[23] D. Ayres-de Campos, C. Y. Spong, E. Chandraharan, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology and Obstetrics*, 131(1):13–24, 2015.

[24] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[25] P. Bakker, G. Colenbrander, A. Verstraeten, and H. Van Geijn. The quality of intrapartum fetal heart rate monitoring. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 116(1):22–27, 2004.

[26] H. D. Banta and S. B. Thacker. Electronic fetal monitoring. *International Journal of Technology Assessment in Health Care*, 18(4):762, 2002.

[27] A. A. Baschat. Ductus venosus Doppler for fetal surveillance in high-risk pregnancies. *Clinical Obstetrics and Gynecology*, 53(4):858–868, 2010.

[28] D. M. Becker, C. A. Tafoya, S. L. Becker, G. H. Kruger, M. J. Tafoya, and T. K. Becker. The use of portable ultrasound devices in low-and middle-income countries: a systematic review of the literature. *Tropical Medicine and International Health*, 21(3):294–311, 2016.

[29] J. Behar, F. Andreotti, S. Zaunseder, Q. Li, J. Oster, and G. D. Clifford. An ECG simulator for generating maternal-foetal activity mixtures on abdominal ECG recordings. *Physiological Measurement*, 35(8):1537–1550, 2014.

[30] J. Behar, F. Andreotti, S. Zaunseder, Q. Li, J. Oster, and G. D. Clifford. An ECG model for simulating maternal-foetal activity mixtures on abdominal ECG recordings. *Physiological Measurements*, 35(8):1537–1550, 2014.

[31] J. Behar, F. Andreotti, S. Zaunseder, J. Oster, and G. D. Clifford. A practical guide to non-invasive foetal electrocardiogram extraction and analysis. *Physiological measurement*, 37(5):R1, 2016.

[32] J. Behar, A. Johnson, G. D. Clifford, and J. Oster. A comparison of single channel fetal ECG extraction methods. *Ann Biomed Eng*, 42(6):1340–1353, jun 2014.

[33] M. Bennasar, J. Martinez, O. Gomez, J. Bartrons, A. Olivella, B. Puerto, and E. Gratacós. Accuracy of four-dimensional spatiotemporal image correlation echocardiography in the prenatal diagnosis of congenital heart defects. *Ultrasound in Obstetrics and Gynecology*, 36(4):458–464, 2010.

[34] A. Bergström, E. Fottrell, H. Hopkins, D. Lloyd, O. Stevenson, and P. Willats. mHealth: can mobile technology improve health in low-and middle-income countries. Technical report, UCL public policy, 2015.

[35] E. Berkley, S. P. Chauhan, A. Abuhamad, and Society for Maternal-Fetal Medicine Publications Committee. Doppler assessment of the fetus with intrauterine growth restriction. *American Journal of Obstetrics and Gynecology*, 206(4):300–308, 2012.

[36] J. Bernardes, A. Costa-Pereira, D. Ayres-de Campos, H. Van Geijn, and L. Pereira-Leite. Evaluation of interobserver agreement of cardiotocograms. *International Journal of Gynecology and Obstetrics*, 57(1):33–37, 1997.

[37] S. C. Blackwell, W. A. Grobman, L. Antoniewicz, M. Hutchinson, and C. G. Bannerman. Interobserver and intraobserver reliability of the NICHD 3-tier

fetal heart rate interpretation system. *American Journal of Obstetrics and Gynecology*, 205(4):378–e1, 2011.

[38] H. Blencowe, S. Cousens, F. B. Jassir, L. Say, D. Chou, C. Mathers, D. Hogan, S. Shiekh, Z. U. Qureshi, D. You, and J. Lawn. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *The Lancet Global Health*, 4(2):e98–e108, 2016.

[39] E. Blix, R. Maude, E. Hals, S. Kisa, E. Karlsen, E. A. Nohr, A. de Jonge, H. Lindgren, S. Downe, L. M. Reinar, and M. Foureur. Intermittent auscultation fetal monitoring during labour: A systematic scoping review to identify methods, effects, and accuracy. *PloS One*, 14(7), 2019.

[40] E. Blix, O. Sviggum, K. S. Koss, and P. Øian. Inter-observer variation in assessment of 845 labour admission tests: comparison between midwives and obstetricians in the clinical setting and two experts. *BJOG: an International Journal of Obstetrics and Gynaecology*, 110(1):1–5, 2003.

[41] E. A. Boamah, K. Asante, K. Ae-Ngibise, P. L. Kinney, D. W. Jack, G. Manu, I. T. Azindow, S. Owusu-Agyei, and B. J. Wylie. Gestational age assessment in the Ghana randomized air pollution and health study (graphs): ultrasound capacity building, fetal biometry protocol development, and ongoing quality control. *JMIR Research Protocols*, 3(4):e77, 2014.

[42] N. Britton, M. A. Miller, S. Safadi, A. Siegel, A. Levine, and M. McCurdy. Tele-ultrasound in resource-limited settings: A systematic review. *Frontiers in Public Health*, 7:244, 2019.

[43] R. Byaruhanga, D. G. Bassani, A. Jagau, P. Muwanguzi, A. L. Montgomery, and J. E. Lawn. Use of wind-up fetal Doppler versus Pinard for fetal heart

rate intermittent monitoring in labour: a randomised clinical trial. *BMJ Open*, 5(1):e006867, 2015.

[44] M. Cannie, J. Jani, S. Dymarkowski, and J. Deprest. Fetal magnetic resonance imaging: luxury or necessity? *Ultrasound in Obstetrics and Gynecology*, 27(5):471–476, 2006.

[45] Cardo Medical. Fetal monitors. monica novii wireless patch system, 2020. Retrieved from `https://cardomedical.com/product-category/fetal-monitors/` on Sep 27 2020.

[46] L. Caserta, Z. Ruggeri, L. D'Emidio, C. Coco, P. Cignini, A. Girgenti, L. Mangiafico, and C. Giorlandino. Two-dimensional fetal echocardiography: where we are. *Journal of Prenatal Medicine*, 2(3):31, 2008.

[47] E. Castillo, D. P. Morales, G. Botella, A. García, L. Parrilla, and A. J. Palma. Efficient wavelet-based ECG processing for single-lead FHR extraction. *Digital Signal Processing*, 23(6):1897–1909, 2013.

[48] Centers for Disease Control and Prevention. PRAMS questionnaires, 2016. Retrieved from `https://www.cdc.gov/prams/questionnaire.htm#questions` on Jun 10 2019.

[49] R. Chaoui, J. Hoffmann, and K. Heling. Three-dimensional (3D) and 4D color Doppler fetal echocardiography using spatio-temporal image correlation (STIC). *Ultrasound in Obstetrics and Gynecology*, 23(6):535–545, 2004.

[50] V. Chudáček, J. Spilka, L. Lhotská, P. Janků, M. Koucký, M. Huptych, and M. Burša. Assessment of features for automatic CTG analysis based on expert annotation. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6051–6054. IEEE, 2011.

[51] S. Clark, P. Sabey, S. Minton, and R. Stoddard. Non-stress testing with acoustic stimulation. In *Seventh Annual Meeting of The Society of Perinatal Obstetricians*, Lake Buena Vista, FL, USA, 1987.

[52] L. Clavier and J. M. Boucher. Segmentation of electrocardiograms using a hidden Markov model. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 1409–1410 vol.4, Oct 1996.

[53] G. Clifford, R. Sameni, J. Ward, J. Robinson, C. Pettigrew, and A. Wolfberg. Clinically accurate fetal ECG parameters acquired from maternal abdominal sensors. *American Journal of Obstetrics and Gynecology*, 201(6):S242, 2009.

[54] G. D. Clifford, I. Silva, J. Behar, and G. B. Moody. Non-invasive fetal ECG analysis. *Physiological Measurement*, 35(8):1521, 2014.

[55] G. D. Clifford and L. Tarassenko. Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. *IEEE transactions on biomedical engineering*, 52(4):630–638, 2005.

[56] G. Cometto, K. Tulenko, A. S. Muula, and R. Krech. Health workforce brain drain: from denouncing the challenge to solving the problem. *PLoS Medicine*, 10(9), 2013.

[57] E. Cosmi, S. Visentin, T. Fanelli, A. J. Mautone, and V. Zanardo. Aortic intima media thickness in fetuses and children with intrauterine growth restriction. *Obstetrics and Gynecology*, 114(5):1109–1114, 2009.

[58] R. Cruz-Martinez, F. Figueras, E. Hernandez-Andrade, D. Oros, and E. Gratacos. Changes in myocardial performance index and aortic isthmus and ductus venosus Doppler in term, small-for-gestational age fetuses with normal umbilical

artery pulsatility index. *Ultrasound in Obstetrics and Gynecology*, 38(4):400–405, 2011.

[59] C. E. V. Cuadros, N. Katebi, F. Marzbanrad, P. Rohloff, and G. D. Clifford. A review of fetal cardiac monitoring, with a focus on low-and middle-income countries. *Physiological Measurement*, 2020.

[60] J. A. Davies, J. A. Spencer, and S. Gallivan. Randomised controlled trial of Doppler ultrasound screening of placental perfusion during pregnancy. *The Lancet*, 340(8831):1299–1303, 1992.

[61] G. Dawes, M. Lobb, M. Moulden, C. Redman, and T. Wheeler. Antenatal cardiotocogram quality and interpretation using computers. *BJOG: An International Journal of Obstetrics and Gynaecology*, 99(10):791–797, 1992.

[62] G. Dawes, G. Visser, J. Goodman, and C. Redman. Numerical analysis of the human fetal heart rate: the quality of ultrasound records. *American Journal of Obstetrics and Gynecology*, 141(1):43–52, 1981.

[63] M. de Onis, M. Blössner, and J. Villar. Levels and patterns of intrauterine growth retardation in developing countries. *European Journal of Clinical Nutrition*, 52 Suppl 1:S5–15, Jan. 1998.

[64] S. Deb, M. S. Mohammed, U. Dhingra, A. Dutta, S. M. Ali, P. Dixit, M. H. Juma, M. J. Hassan, S. Sazawal, I. Nisar, and M. Ilyas. Performance of late pregnancy biometry for gestational age dating in low-income and middle-income countries: a prospective, multicountry, population-based cohort study from the who alliance for maternal and newborn health improvement (AMANHI) study group. *The Lancet Global Health*, 8(4):e545–e554, 2020.

[65] J. Deng, J. E. Gardener, C. H. Rodeck, and W. R. Lees. Fetal echocardiography

in three and four dimensions. *Ultrasound in Medicine and Biology*, 22(8):979–986, 1996.

[66] D. Devane, J. G. Lalor, S. Daly, W. McGuire, A. Cuthbert, and V. Smith. Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing. *Cochrane Database of Systematic Reviews*, (1), 2017.

[67] G. R. DeVore, J. Horenstein, B. Siassi, and L. D. Platt. Fetal echocardiography: VII. Doppler color flow mapping: A new technique for the diagnosis of congenital heart disease. *American Journal of Obstetrics and Gynecology*, 156(5):1054–1064, 1987.

[68] P. M. Dietz, L. J. England, W. M. Callaghan, M. Pearl, M. L. Wier, and M. Kharrazi. A comparison of LMP-based and ultrasound-based estimates of gestational age using linked California livebirth and prenatal screening records. *Paediatric and Perinatal Epidemiology*, 21(s2):62–71, 2007.

[69] A. Dražančić. Antenatal care in developing countries. What should be done? *Journal of Perinatal Medicine*, 29(3):188–198, 2001.

[70] C. E. East, L. Begg, P. B. Colditz, and R. Lau. Fetal pulse oximetry for fetal assessment in labour. *Cochrane Database of Systematic Reviews*, (10), 2014.

[71] R. D. Eden, F. H. Boehm, and M. Haire. *Assessment and Care of the Fetus: Physiological, clinical, and medicolegal principles*. Appleton & Lange, Norwalk, Conn, 1990.

[72] D. H. Evans. *Doppler ultrasound: Physics, instrumentation, and clinical applications*. John Wiley & Sons, Hoboken, NJ, 1989.

[73] G. Eysenbach and Consort-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *Journal of Medical Internet Research*, 13(4):e126, 2011.

[74] K. L. Fernando, V. J. Mathews, M. W. Varner, and E. B. Clark. Robust estimation of fetal heart rate variability using doppler ultrasound. *IEEE transactions on biomedical engineering*, 50(8):950–957, 2003.

[75] A. Feroz, S. Perveen, and W. Aftab. Role of mhealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. *BMC Health Services Research*, 17(1):704, 2017.

[76] F. Figueras and J. Gardosi. Intrauterine growth restriction: new concepts in antenatal surveillance, diagnosis, and management. *American Journal of Obstetrics and Gynecology*, 204(4):288–300, 2011.

[77] K. Finlayson and S. Downe. Why do women not use antenatal services in low- and middle-income countries? A meta-synthesis of qualitative studies. *PLoS Medicine*, 10(1):e1001373, 2013.

[78] V. Flenady, P. Middleton, G. C. Smith, W. Duke, J. J. Erwich, T. Y. Khong, J. Neilson, M. Ezzati, L. Koopmans, D. Ellwood, and R. Fretts. Stillbirths: the way forward in high-income countries. *The Lancet*, 377(9778):1703–1717, 2011.

[79] A. Fort, L. Masotti, S. Rocchi, V. Vignoli, M. Di Tommaso, C. Consoli, and F. Branconi. Parametric spectral analysis of umbilical artery doppler signals for fetal heart rate variability evaluation. *European journal of ultrasound*, 2(4):321–328, 1995.

[80] H. L. Franklin, W. Mirza, D. L. Swanson, J. E. Newman, R. L. Goldenberg, D. Muyodi, L. Figueroa, R. O. Nathan, J. O. Swanson, N. Goldsmith, and

N. Kanaiza. Factors influencing referrals for ultrasound-diagnosed complications during prenatal care in five low and middle income countries. *Reproductive Health*, 15(1):204, 2018.

[81] R. K. Freeman, G. Anderson, and W. Dorchester. A prospective multi-institutional study of antepartum fetal heart rate monitoring: I. Risk of perinatal mortality and morbidity according to antepartum fetal heart rate test results. *American Journal of Obstetrics and Gynecology*, 143(7):771–777, 1982.

[82] R. K. Freeman, T. J. Garite, M. P. Nageotte, and L. A. Miller. *Fetal heart rate monitoring*. Lippincott Williams & Wilkins, Philadelphia, PA, 2012.

[83] H. L. Galan, E. Ferrazzi, and J. C. Hobbins. Intrauterine growth restriction (IUGR): biometric and Doppler assessment. *Prenatal Diagnosis: Published in Affiliation with the International Society for Prenatal Diagnosis*, 22(4):331–337, 2002.

[84] K. B. Gan, E. Zahedi, and M. A. M. Ali. Transabdominal fetal heart rate detection using NIR photopleythysmography: instrumentation and clinical results. *IEEE Transactions on Biomedical Engineering*, 56(8):2075–2082, 2009.

[85] W. Ganzevoort, N. Mensing Van Charante, B. Thilaganathan, F. Prefumo, B. Arabin, C. M. Bilardo, C. Brezinka, J. Derks, A. Diemert, J. J. Duvekot, and E. Ferrazzi. How to monitor pregnancies complicated by fetal growth restriction and delivery before 32 weeks: post-hoc analysis of truffle study. *Ultrasound in Obstetrics and Gynecology*, 49(6):769–777, 2017.

[86] A. Georgieva, P. Abry, V. Chudáček, P. M. Djurić, M. G. Frasch, R. Kok, C. A. Lear, S. N. Lemmens, I. Nunes, A. T. Papageorghiou, et al. Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd workshop on

signal processing and monitoring in labor (October 2017, Oxford, UK). *Acta Obstetricia et Gynecologica Scandinavica*, 98(9):1207–1217, 2019.

[87] W. Giles, A. Bisits, S. O'Callaghan, and A. Gill. The Doppler Assessment in multiple pregnancy randomised controlled trial of ultrasound biometry versus umbilical artery Doppler ultrasound and biometry in twin pregnancy. *BJOG: An International Journal of Obstetrics and Gynaecology*, 110(6):593–597, 2003.

[88] A. C. Gittenberger-de Groot, M. R. Jongbloed, M. C. de Ruiter, M. Bartelings, and P. R. E. Cardiac morphogenesis. In S. Yagel, N. H. Silverman, and U. Gembruch, editors, *Fetal Cardiology*, chapter 1, pages 1–18. Taylor & Francis Group, Boca Raton, FL, 2019.

[89] M. Godfrey, B. Messing, S. Cohen, D. Valsky, and S. Yagel. Functional assessment of the fetal heart: A review. *Ultrasound in Obstetrics and Gynecology*, 39(2):131–144, 2012.

[90] R. L. Goldenberg, E. M. McClure, and C. M. Bann. The relationship of intrapartum and antepartum stillbirth rates to measures of obstetric care in developed and developing countries. *Acta Obstetricia et Gynecologica Scandinavica*, 86(11):1303–1309, 2007.

[91] R. L. Goldenberg, E. M. McClure, S. Saleem, and U. M. Reddy. Infection-related stillbirths. *The Lancet*, 375(9724):1482–1490, 2010.

[92] R. L. Goldenberg, R. O. Nathan, D. Swanson, S. Saleem, W. Mirza, F. Esamai, D. Muyodi, A. L. Garces, L. Figueroa, E. Chomba, and M. Chiwala. Routine antenatal ultrasound in low-and middle-income countries: first look–a cluster randomised trial. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(12):1591–1599, 2018.

[93] R. L. Goldenberg, S. Saleem, O. Pasha, M. S. Harrison, and E. M. Mcclure. Reducing stillbirths in low-income countries. *Acta Obstetricia et Gynecologica Scandinavica*, 95(2):135–143, 2016.

[94] M. D. Gomez-Roig, E. Mazarico, E. Valladares, L. Guirado, M. Fernandez-Arias, and A. Vela. Aortic intima-media thickness and aortic diameter in small for gestational age and growth restricted fetuses. *PLoS One*, 10(5):e0126842, 2015.

[95] W. Graham, S. Woodd, P. Byass, V. Filippi, G. Gon, S. Virgo, D. Chou, S. Hounton, R. Lozano, R. Pattinson, et al. Diversity and divergence: the dynamic burden of poor maternal health. *The Lancet*, 388(10056):2164–2175, 2016.

[96] S. Graja and J. Boucher. Multiscale hidden Markov model applied to ECG segmentation. In *IEEE International Symposium on Intelligent Signal Processing, 2003*, pages 105–109, Sep. 2003.

[97] N. Greenwold, S. Wallace, A. Prost, and E. Jauniaux. Implementing an obstetric ultrasound training program in rural Africa. *International Journal of Gynecology and Obstetrics*, 124(3):274–277, 2014.

[98] R. M. Grivell, Z. Alfirevic, G. M. Gyte, and D. Devane. Antenatal cardiotocography for fetal assessment. *Cochrane Database of Systematic Reviews*, (9), 2015.

[99] R. S. Groen, J. J. Leow, V. Sadasivam, and A. L. Kushner. Indications for ultrasound use in low-and middle-income countries. *Tropical Medicine and International Health*, 16(12):1525–1535, 2011.

[100] P. Hamelmann, R. Vullings, A. F. Kolen, J. W. Bergmans, J. O. van Laar, P. Tortoli, and M. Mischi. Doppler ultrasound technology for fetal heart rate

monitoring: a review. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(2):226–238, 2019.

[101] Y. W. Han, Y. Fardini, C. Chen, K. G. Iacampo, V. A. Peraino, J. M. Shamonki, and R. W. Redline. Term stillbirth caused by oral fusobacterium nucleatum. *Obstetrics and Gynecology*, 115(2 Pt 2):442, 2010.

[102] M. J. Hasan, M. S. Hossain, Y. Arafat, G. M. R. Karim, Z. Ahmed, and M. S. Mafi. Mobile application can be an effective tool for reduction of maternal mortality. *International Journal of Perceptions in Public Health*, 1(2):92–94, 2017.

[103] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[104] J. I. Hoffman and R. Christianson. Congenital heart disease in a cohort of 19,502 births with long-term follow-up. *The American Journal of Cardiology*, 42(4):641–647, 1978.

[105] E. Hon. The instrumentation of fetal heart rate and fetal electrocardiography. I. A fetal heart monitor. *Connecticut Medicine*, 24:289, 1960.

[106] E. Hon and O. Hess. The clinical value of fetal electrocardiography. *American Journal of Obstetrics and Gynecology*, 79:1012–1023, 1960.

[107] E. Hon and S. Lee. Averaging techniques in fetal electrocardiography. *Medical Electronics and Biological Engineering*, 2(1):71–76, 1964.

[108] N. Housseine, M. C. Punt, J. L. Browne, T. Meguid, K. Klipstein-Grobusch, B. E. Kwast, A. Franx, D. E. Grobbee, and M. J. Rijken. Strategies for intrapartum foetal surveillance in low-and middle-income countries: A systematic review. *PloS One*, 13(10):e0206295, 2018.

[109] N. Housseine, M. C. Punt, J. L. Browne, J. van't Hooft, N. Maaløe, T. Meguid, G. B. Theron, A. Franx, D. E. Grobbee, G. H. Visser, et al. Delphi consensus statement on intrapartum fetal monitoring in low-resource settings. *International Journal of Gynecology and Obstetrics*, 146(1):8–16, 2019.

[110] D. Hoyer, F. Tetschke, S. Jaekel, S. Nowack, O. W. Witte, E. Schleußner, and U. Schneider. Fetal functional brain age assessed from universal developmental indices obtained from neuro-vegetative activity patterns. *PloS One*, 8(9):e74431, 2013.

[111] D. Hoyer, J. Żebrowski, D. Cysarz, H. Gonçalves, A. Pytlik, C. Amorim-Costa, J. Bernardes, D. Ayres-de Campos, O. W. Witte, E. Schleußner, and L. Stroux. Monitoring fetal maturation—objectives, techniques and indices of autonomic function. *Physiological Measurement*, 38(5):R61, 2017.

[112] L. Hruban, J. Spilka, V. Chudáček, P. Janků, M. Huptych, Michal nd Burša, A. Hudec, M. Kacerovskỳ, M. Kouckỳ, M. Procházka, and V. Korečko. Agreement on intrapartum cardiotocogram recordings between expert obstetricians. *Journal of Evaluation in Clinical Practice*, 21(4):694–702, 2015.

[113] N. P. Hughes. *Probabilistic models for automated ECG interval analysis*. PhD thesis, University of Oxford, 2006.

[114] N. P. Hughes, L. Tarassenko, and S. J. Roberts. Markov models for automated ECG interval analysis. *Advances in Neural Information Processing Systems*, 16:611–618, 2004.

[115] C. Ionescu. The benefits of 3D-4D fetal echocardiography. *Maedica*, 5(1):45, 2010.

[116] R. Jaros, R. Martinek, and R. Kahankova. Non-adaptive methods for fetal ECG signal processing: A review and appraisal. *Sensors*, 18(11):3648, 2018.

[117] W. Jatmiko, M. A. Ma'Sum, S. M. Isa, E. Imah, R. Rahmatullah, and B. Wiweko. Developing smart telehealth system in Indonesia: Progress and challenge. In *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 29–36. IEEE, 2015.

[118] E. Jauniaux and F. Prefumo. Fetal heart monitoring in labour: from Pinard to artificial intelligence. *BJOG-An International Journal of Obstetrics and Gynaecology*, 123(6):870, 2016.

[119] J. Jezewski, A. Matonia, T. Kupka, D. Roj, and R. Czabanski. Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram. *Biomedizinische Technik/Biomedical Engineering*, 57(5):383–394, 2012.

[120] J. Jezewski, D. Roj, J. Wrobel, and K. Horoba. A novel technique for fetal heart rate estimation from Doppler ultrasound signal. *Biomedical Engineering Online*, 10(1):92, 2011.

[121] J. Jezewski, J. Wrobel, A. Matonia, K. Horoba, R. Martinek, T. Kupka, and M. Jezewski. Is abdominal fetal electrocardiography an alternative to Doppler ultrasound for FHR variability evaluation? *Frontiers in Physiology*, 8:305, 2017.

[122] M. Juarez, Y. Juarez, E. Coyote, T. Nguyen, C. Shaw, R. Hall-Clifford, G. Clifford, and P. Rohloff. Working with lay midwives to improve the detection of neonatal complications in rural guatemala. *BMJ Open Quality*, 9(1), 2020.

[123] R. Kahankova, R. Martinek, R. Jaros, K. Behbehani, A. Matonia, M. Jezewski, and J. A. Behar. A review of signal processing techniques for non-invasive fetal electrocardiography. *IEEE Reviews in Biomedical Engineering*, 13:51–73, 2019.

[124] C. Kähler, B. Grimm, E. Schleussner, A. Schneider, U. Schneider, H. Nowak, L. Vogt, and H.-J. Seewald. The application of fetal magnetocardiography

(FMCG) to investigate fetal arrhythmias and congenital heart defects (CHD). *Prenatal Diagnosis: Published in Affiliation with the International Society for Prenatal Diagnosis*, 21(3):176–182, 2001.

[125] B. A. Kamala, H. L. Ersdal, I. Dalen, M. S. Abeid, M. M. Ngarina, J. M. Perlman, and H. L. Kidanto. Implementation of a novel continuous fetal Doppler (Moyo) improves quality of intrapartum fetal heart rate monitoring in a resource-limited tertiary hospital in tanzania: An observational study. *PLoS One*, 13(10):e0205698, 2018.

[126] B. A. Kamala, H. L. Ersdal, I. Dalen, M. S. Abeid, M. M. Ngarina, J. M. Perlman, and H. L. Kidanto. Implementation of a novel continuous fetal Doppler (Moyo) improves quality of intrapartum fetal heart rate monitoring in a resource-limited tertiary hospital in Tanzania: An observational study. *PLoS ONE*, 13(10):1–14, 2018.

[127] B. A. Kamala, H. L. Kidanto, P. J. Wangwe, I. Dalen, E. R. Mduma, J. M. Perlman, and H. L. Ersdal. Intrapartum fetal heart rate monitoring using a handheld Doppler versus Pinard stethoscope: a randomized controlled study in Dar Es Salaam. *International Journal of Women's Health*, 10:341, 2018.

[128] H. Kapaya, R. Jacques, and D. Anumba. Comparison of diurnal variations, gestational age and gender related differences in fetal heart rate (FHR) parameters between appropriate-for-gestational-age (AGA) and small-for-gestational-age (SGA) fetuses in the home environment. *PloS One*, 13(3), 2018.

[129] V. Kariniemi and K. Hukkinen. Quantification of fetal heart rate variability by magnetocardiography and direct electrocardiography. *American journal of obstetrics and gynecology*, 128(5):526–530, 1977.

[130] N. Katebi, F. Marzbanrad, L. Stroux, C. E. Valderrama, and G. D. Clifford.

Unsupervised hidden semi-Markov model for automatic beat onset detection in 1D Doppler ultrasound. *Physiological Measurement*, 41(8):085007, 2020.

[131] N. Katebi, R. Sameni, and G. D. Clifford. Deep Sequence Learning for Accurate Gestational Age Estimation from a $25 Doppler Device. *ML4H, NeurIPS*, 2020.

[132] N. H. Khan, E. Tegnander, J. M. Dreier, S. Eik-Nes, H. Torp, and G. Kiss. Automatic measurement of the fetal abdominal section on a portable ultrasound machine for use in low and middle income countries. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2016.

[133] I. Kiefer-Schmidt, M. Lim, A. Wacker-Gußmann, E. Ortiz, H. Abele, K. O. Kagan, R. Kaulitz, D. Wallwiener, and H. Preissl. Fetal magnetocardiography (FMCG): Moving forward in the establishment of clinical reference data by advanced biomagnetic instrumentation and analysis. *Journal of Perinatal Medicine*, 40(3):277–286, 2012.

[134] E. T. Kim, K. Singh, A. Moran, D. Armbruster, and N. Kozuki. Obstetric ultrasound use in low and middle income countries: a narrative review. *Reproductive Health*, 15(1):129, 2018.

[135] H. H. Kimberly, A. Murray, M. Mennicke, A. Liteplo, J. Lew, J. S. Bohan, L. Tyer-Viola, R. Ahn, T. Burke, and V. E. Noble. Focused maternal ultrasound by midwives in rural Zambia. *Ultrasound in Medicine an Biology*, 36(8):1267–1272, 2010.

[136] T. R. Kohler and D. S. Sumner. Vascular laboratory: Arterial physiologic assessment. In J. L. Cronenwett and K. W. Johnston, editors, *Rutherford's Vascular Surgery*, chapter 15, pages 214–229. Elsevier Health Sciences, Philadelphia, PA, 2014.

[137] J. Kohlmorgen and S. Lemm. A dynamic HMM for on-line segmentation of sequential data. *Advances in Neural Information Processing Systems*, pages 793–800, 2002.

[138] A. Koski. Modelling ECG signals with hidden Markov models. *Artificial Intelligence in Medicine*, 8(5):453–471, 1996.

[139] F. Kovács, C. Horváth, Á. T. Balogh, and G. Hosszú. Fetal phonocardiography-past and future possibilities. *Computer Methods and Programs in Biomedicine*, 104(1):19–25, 2011.

[140] N. Kozuki, L. C. Mullany, S. K. Khatry, R. K. Ghimire, S. Paudel, K. Blakemore, C. Bird, J. M. Tielsch, S. C. LeClerq, and J. Katz. Accuracy of home-based ultrasonographic diagnosis of obstetric risk factors by primary-level health workers in rural Nepal. *Obstetrics and Gynecology*, 128(3):604, 2016.

[141] S. R. Lafontan, H. L. Kidanto, H. L. Ersdal, C. K. Mbekenga, and J. Sundby. Perceptions and experiences of skilled birth attendants on using a newly developed strap-on electronic fetal heart rate monitor in Tanzania. *BMC Pregnancy and Childbirth*, 19(1):165, 2019.

[142] K.-C. Lai and J. J. Shynk. A successive cancellation algorithm for fetal heart-rate estimation using an intrauterine ECG signal. *IEEE Transactions on Biomedical Engineering*, 49(9):943–954, 2002.

[143] I. Lakhno. Autonomic imbalance captures maternal and fetal circulatory response to pre-eclampsia. *Clinical Hypertension*, 23(1):5, 2017.

[144] S. Larks and G. Larks. Components of the fetal electrocardiogram and intrauterine electrical axis: Quantitative data. *Neonatology*, 10(3-4):140–152, 1966.

[145] S. D. Larks and L. D. Longo. Electrocardiographic studies of the fetal heart during delivery. *Obstetrics and Gynecology*, 19(6):740–747, 1962.

[146] J. E. Lawn, H. Blencowe, S. Oza, D. You, A. C. C. Lee, P. Waiswa, M. Lalli, Z. Bhutta, A. J. D. Barros, P. Christian, C. Mathers, and S. N. Cousens. Every newborn: Progress, priorities, and potential beyond survival. *Lancet*, 384(9938):189–205, July 2014.

[147] J. E. Lawn, H. Blencowe, S. Oza, D. You, A. C. C. Lee, P. Waiswa, M. Lalli, Z. Bhutta, A. J. D. Barros, P. Christian, C. Mathers, and S. N. Cousens. Every Newborn: progress, priorities, and potential beyond survival. *Lancet*, 384(9938):189–205, jul 2014.

[148] J. E. Lawn, H. Blencowe, P. Waiswa, A. Amouzou, C. Mathers, D. Hogan, V. Flenady, J. F. Frøen, Z. U. Qureshi, C. Calderwood, and S. Shiekh. Stillbirths: rates, risk factors, and acceleration towards 2030. *The Lancet*, 387(10018):587–603, 2016.

[149] J. E. Lawn, A. C. Lee, M. Kinney, L. Sibley, W. A. Carlo, V. K. Paul, R. Pattinson, and G. L. Darmstadt. Two million intrapartum-related stillbirths and neonatal deaths: where, why, and what can be done? *International Journal of Gynecology and Obstetrics*, 107(Supplement):S5–S19, 2009.

[150] J. E. Lawn, A. C. Lee, M. Kinney, L. Sibley, W. A. Carlo, V. K. Paul, R. Pattinson, and G. L. Darmstadt. Two million intrapartum-related stillbirths and neonatal deaths: where, why, and what can be done? *International Journal of Gynecology & Obstetrics*, 107(Supplement):S5–S19, 2009.

[151] A. C. Lee, J. Katz, H. Blencowe, S. Cousens, N. Kozuki, J. P. Vogel, L. Adair, A. H. Baqui, Z. A. Bhutta, L. E. Caulfield, and P. Christian. National and regional estimates of term and preterm babies born small for gestational age in

138 low-income and middle-income countries in 2010. *The Lancet Global Health*, 1(1):e26–e36, 2013.

[152] C. C. Lees, N. Marlow, A. van Wassenaer-Leemhuis, B. Arabin, C. M. Bilardo, C. Brezinka, S. Calvert, J. B. Derks, A. Diemert, J. J. Duvekot, and E. Ferrazzi. 2 year neurodevelopmental and intermediate perinatal outcomes in infants with very preterm fetal growth restriction (TRUFFLE): a randomised trial. *The Lancet*, 385(9983):2162–2172, 2015.

[153] Y. Leitner, A. Fattal-Valevski, R. Geva, H. Bassan, E. Posner, M. Kutai, A. Many, A. J. Jaffa, and S. Harel. Six-year follow-up of children with intrauterine growth retardation: long-term, prospective study. *Journal of Child Neurology*, 15(12):781–786, 2000.

[154] R. Liston, D. Sawchuck, D. Young, N. Brassard, K. Campbell, G. Davies, W. Ehman, D. Farine, D. Farquharson, E. Hamilton, and M. Helewa. Fetal health surveillance: Antepartum and intrapartum consensus guideline. *Journal of Obstetrics and Gynaecology Canada*, 29(9):S3–S4, 2007.

[155] A. D. Lopez, C. AbouZahr, K. Shibuya, and L. Gollogly. Keeping count: Births, deaths, and causes of death. *The Lancet*, 370(9601):1744–1746, 2007.

[156] G. Magenes, M. Signorini, and R. Sassi. Automatic diagnosis of fetal heart rate: comparison of different methodological approaches. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 2, pages 1604–1607. IEEE, 2001.

[157] J. Mahdizadeh, H. Bouraghi, S. S. G. S. Panahi, A. Mohammadpour, A. K. Shargh, M. R. Mojarad, and M. Kahouei. A theory map of the causes of perinatal death in a developing country. *Crescent Journal of Medical and Biological Sciences*, 6(2):237–241, 2019.

[158] K. Mahomed, R. Nyoni, T. Mulambo, J. Kasule, and E. Jacobus. Randomised controlled trial of intrapartum fetal heart rate monitoring. *BMJ*, 308(6927):497–500, 1994.

[159] K. Mäkikallio, J. Räsänen, T. Mäkikallio, O. Vuolteenaho, and J. Huhta. Human fetal cardiovascular profile score and neonatal outcome in intrauterine growth restriction. *Ultrasound in Obstetrics and Gynecology*, 31(1):48–54, 2008.

[160] N. Malhotra, J. Malhotra, V. Mathur, S. Tomar, K. Singh, J. Rao, S. Gupta, and N. Malhotra. Antenatal assessment of fetal well-being. In N. Malhotra, P. Kumar, S. Panchal, K. Shah, P. Acharya, and J. Malhotra, editors, *Ultrasound in Obstetrics and Gynecology*, chapter 26, pages 227–241. JP Medical Ltd, London, UK, 2014.

[161] N. Manwani and P. Sastry. Noise tolerance under risk minimization. *IEEE Transactions on Cybernetics*, 43(3):1146–1151, 2013.

[162] G. Mari and F. Hanif. Fetal Doppler: umbilical artery, middle cerebral artery, and venous system. *Seminars in Perinatology*, 32(4):253–257, 2008.

[163] E. N. Marieb and K. Hoehn. *Human Anatomy & Physiology*. Pearson Education, Boston, MA, USA, 2007.

[164] S. M. Martens, C. Rabotti, M. Mischi, and R. J. Sluijter. A robust fetal ECG detection method for abdominal recordings. *Physiological Measurement*, 28(4):373, 2007.

[165] B. Martinez, E. Coyote, R. Hall-Clifford, M. Juarez, A. C. Miller, A. Francis, C. E. Valderrama, L. Stroux, G. D. Clifford, and P. Rohloff. mHealth intervention to improve the continuum of maternal and perinatal care in rural Guatemala: a pragmatic, randomized controlled feasibility trial. *Reproductive Health*, 15(1):120, 2018.

[166] B. Martinez, R. Hall-Clifford, E. Coyote, L. Stroux, C. E. Valderrama, C. Aaron, A. Francis, C. Hendren, P. Rohloff, and G. D. Clifford. Agile development of a smartphone app for perinatal monitoring in a resource-constrained setting. *Journal of Health Informatics in Developing Countries*, 11(1), 2017.

[167] B. Martinez, E. C. Ixen, R. Hall-Clifford, M. Juarez, A. C. Miller, A. Francis, C. E. Valderrama, L. Stroux, G. D. Clifford, and P. Rohloff. mhealth intervention to improve the continuum of maternal and perinatal care in rural guatemala: a pragmatic, randomized controlled feasibility trial. *Reproductive Health*, 15(1):120, 2018.

[168] R. Martis, O. Emilia, D. S. Nurdiati, and J. Brown. Intermittent auscultation (IA) of fetal heart rate in labour for fetal well-being. *Cochrane Database of Systematic Reviews*, (2), 2017.

[169] F. Marzbanrad. *Modeling fetal cardiac valve intervals and fetal-maternal interactions*. PhD thesis, The University of Melbourne, 2015.

[170] F. Marzbanrad, A. H. Khandoker, Y. Kimura, M. Palaniswami, and G. D. Clifford. Estimating fetal gestational age using cardiac valve intervals. In *2016 Computing in Cardiology Conference (CinC)*, pages 109–112. IEEE, 2016.

[171] F. Marzbanrad, A. H. Khandoker, Y. Kimura, M. Palaniswami, and G. D. Clifford. Assessment of fetal development using cardiac valve intervals. *Frontiers in Physiology*, 8:313, 2017.

[172] F. Marzbanrad, Y. Kimura, K. Funamoto, S. Oshio, M. Endo, N. Sato, M. Palaniswami, and A. H. Khandoker. Model-based estimation of aortic and mitral valves opening and closing timings in developing human fetuses. *IEEE Journal of Biomedical and Health Informatics*, 20(1):240–248, 2014.

[173] F. Marzbanrad, Y. Kimura, K. Funamoto, R. Sugibayashi, M. Endo, T. Ito, M. Palaniswami, and A. H. Khandoker. Automated estimation of fetal cardiac timing events from Doppler ultrasound signal using hybrid models. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1169–1177, 2013.

[174] F. Marzbanrad, L. Stroux, and G. D. Clifford. Cardiotocography and beyond: a review of one-dimensional Doppler ultrasound application in fetal monitoring. *Physiological Measurement*, 39(8):08TR01, 2018.

[175] P. F. Mdoe, H. L. Ersdal, E. Mduma, R. Moshiro, I. Dalen, J. M. Perlman, and H. Kidanto. Randomized controlled trial of continuous Doppler versus intermittent fetoscope fetal heart rate monitoring in a low-resource setting. *International Journal of Gynecology and Obstetrics*, 143(3):344–350, 2018.

[176] P. F. Mdoe, H. L. Ersdal, E. R. Mduma, J. M. Perlman, R. Moshiro, P. T. Wangwe, and H. Kidanto. Intermittent fetal heart rate monitoring using a fetoscope or hand held Doppler in rural Tanzania: A randomized controlled trial. *BMC Pregnancy and Childbirth*, 18(1):1–8, 2018.

[177] F. C. Miller, K. E. Pearse, and R. H. Paul. Fetal heart rate pattern recognition by the method of auscultation. *Obstetrics and Gynecology*, 64(3):332–336, 1984.

[178] F. Mone, F. M. McAuliffe, and S. Ong. The clinical application of Doppler ultrasound in obstetrics. *The Obstetrician and Gynaecologist*, 17(1):13–19, 2015.

[179] M. Morgenstern, J. D. Sargent, and R. Hanewinkel. Relation between socioeconomic status and body mass index: evidence of an indirect path via television use. *Archives of Pediatrics and Adolescent Medicine*, 163(8):731–738, 2009.

[180] R. Morris, R. Say, S. Robson, J. Kleijnen, and K. Khan. Systematic review and meta-analysis of middle cerebral artery Doppler to predict perinatal wellbe-

ing. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 165(2):141–155, 2012.

[181] S. Munabi-Babigumira, C. Glenton, M. Willcox, and H. Nabudere. Ugandan health workers' and mothers' views and experiences of the quality of maternity care and the use of informal solutions: A qualitative study. *PloS One*, 14(3):e0213511, 2019.

[182] M. K. Mwaniki, M. Atieno, J. E. Lawn, and C. R. Newton. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *The Lancet*, 379(9814):445–452, 2012.

[183] J. P. Neilson. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database of Systematic Reviews*, (4), 1998.

[184] K. Nicolaides, G. Rizzo, K. Hecher, and R. Ximenes. *Doppler in obstetrics*. The Fetal Medicine Foundation, London, UK, 2002.

[185] Y. Nomura and T. Kurita. Robust training of deep neural networks with noisy labels by graph label propagation. In *International Workshop on Frontiers of Computer Vision*, pages 281–293. Springer, 2021.

[186] H. Norén, I. Amer-Wåhlin, H. Hagberg, A. Herbst, I. Kjellmer, K. Marşál, P. Olofsson, and K. G. Rosén. Fetal electrocardiogram in labor and neonatal outcome: Data from the Swedish randomized controlled trial on intrapartum fetal monitoring. *American Journal of Obstetrics and Gynecology*, 188:183–192, 2003.

[187] H. Norén, S. Blad, A. Carlsson, A. Flisberg, A. Gustavsson, H. Lilja, M. Wennergren, and H. Hagberg. STAN in clinical practice? The outcome of 2 years of regular use in the city of gothenburg. *American Journal of Obstetrics and Gynecology*, 195(1):7–15, 2006.

[188] M. Ome-Kaius, S. Karl, R. A. Wangnapi, J. W. Bolnga, G. Mola, J. Walker, I. Mueller, H. W. Unger, and S. J. Rogerson. Effects of plasmodium falciparum infection on umbilical artery resistance and intrafetal blood flow distribution: a Doppler ultrasound study from Papua New Guinea. *Malaria Journal*, 16(1):35, 2017.

[189] W. H. Organization et al. *WHO compendium of innovative health technologies for low resource settings, 2011-2014: assistive devices, eHealth solutions, medical devices, other technologies, technologies for outbreaks*. World Health Organization, 2015.

[190] V. O'Dwyer, G. Burke, J. Unterscheider, S. Daly, M. P. Geary, M. M. Kennelly, F. M. McAuliffe, K. O'Donoghue, A. Hunter, J. J. Morrison, and P. Dicker. Defining the residual risk of adverse perinatal outcome in growth-restricted fetuses with normal umbilical artery blood flow. *American Journal of Obstetrics and Gynecology*, 211(4):420–e1, 2014.

[191] Z. Papp and T. Fekete. The evolving role of ultrasound in obstetrics/gynecology practice. *International Journal of Gynecology and Obstetrics*, 82(3):339–346, 2003.

[192] J. Pardey, M. Moulden, and C. W. Redman. A computer system for the numerical analysis of nonstress tests. *American Journal of Obstetrics and Gynecology*, 186(5):1095–1103, 2002.

[193] Y. Park, K. Lee, D. Youn, N. Kim, W. Kim, and S. Park. On detecting the presence of fetal R-wave using the moving averaged magnitude difference algorithm. *IEEE Transactions on Biomedical Engineering*, 39(8):868–871, 1992.

[194] G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1944–1952, 2017.

[195] R. Pattinson, K. Kerber, P. Waiswa, L. T. Day, F. Mussell, S. Asiruddin, H. Blencowe, and J. E. Lawn. Perinatal mortality audit: Counting, accountability, and overcoming challenges in scaling up in low-and middle-income countries. *International Journal of Gynecology and Obstetrics*, 107(Supplement):S113–S122, 2009.

[196] M. Peters, J. Crowe, J. F. Piéri, H. Quartero, B. Hayes-Gill, D. James, J. Stinstra, and S. Shakespeare. Monitoring the fetal heart non-invasively: a review of methods. *Journal of Perinatal Medical*, 29(5):408–416, 2001.

[197] S. Pildner von Steinburg, A. Boulesteix, C. Lederer, S. Grunow, S. Schiermeier, W. Hatzmann, K. M. Schneider, and M. Daumer. What is the "normal" fetal heart rate? *PeerJ*, 1:e82, 2013.

[198] M. Plotkin, B. Kamala, J. Ricca, L. Fogarty, S. Currie, H. Kidanto, and S. B. Wheeler. Systematic review of Doppler for detecting intrapartum fetal heart abnormalities and measuring perinatal mortality in low-and middle-income countries. *International Journal of Gynecology and Obstetrics*, 2019.

[199] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[200] H. Rahman, P. Renjhen, and S. Dutta. Reliability of admission cardiotocography for intrapartum monitoring in low resource setting. *Nigerian Medical Journal: Journal of the Nigeria Medical Association*, 53(3):145, 2012.

[201] F. Rebelo, D. R. Farias, R. H. Mendes, M. M. Schlüssel, and G. Kac. Blood pressure variation throughout pregnancy according to early gestational BMI: a Brazilian cohort. *Arquivos Brasileiros de Cardiologia*, 104(4):284–291, 2015.

[202] E. A. Reece, S. Gabrielli, N. Degennaro, and J. C. Hobbins. Dating through pregnancy: A measure of growing up. *Obstetrical and Gynecological Survey*, 44(7):544–555, 1989.

[203] M. Ren, W. Zeng, B. Yang, and R. Urtasun. Learning to reweight examples for robust deep learning. In *International Conference on Machine Learning*, pages 4334–4343. PMLR, 2018.

[204] I. Rezek and S. J. Roberts. Envelope extraction via complex homomorphic filtering. *Technical Report TR-98-9 Technical report*, 1998.

[205] M. Romano, L. Iuppariello, A. M. Ponsiglione, G. Improta, P. Bifulco, and M. Cesarelli. Frequency and time domain analysis of foetal heart rate variability with traditional indexes: a critical survey. *Computational and Mathematical Methods in Medicine*, 2016, 2016.

[206] Royal College of Obstetricians and Gynaecologists. Small-for-Gestational-Age Fetus, Investigation and Management (Green-top Guideline No. 31), 2002.

[207] R. Sameni and G. D. Clifford. A review of fetal ECG signal processing; issues and promising directions. *The Open Pacing, Electrophy and Therapy Journal*, 3:4–20, 2010.

[208] R. Sameni and G. D. Clifford. A review of fetal ECG signal processing; issues and promising directions. *The Open Pacing, Electrophysiology & Therapy Journal*, 3:4, 2010.

[209] R. Sameni, C. Jutten, and M. B. Shamsollahi. What ICA provides for ECG processing: Application to noninvasive fetal ECG extraction. In *2006 IEEE International Symposium on Signal Processing and Information Technology*, pages 656–661, Vancouver, BC, 2006.

[210] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford. A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Transactions on Biomedical Engineering*, 54(12):2172–2185, 2007.

[211] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford. A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Transactions on Biomedical Engineering*, 54(12):2172–2185, 2007.

[212] T. P. Sartwelle, J. C. Johnston, and B. Arda. A half century of electronic fetal monitoring and bioethics: silence speaks louder than words. *Maternal Health, Neonatology and Perinatology*, 3(1):21, Nov 2017.

[213] Save the Children. State of the world's newborns. Technical report, World Health Organization, Washington D.C., 2001.

[214] A. Schmidt, R. Witte, L. Swiderski, J. Zöllkau, U. Schneider, and D. Hoyer. Advanced automatic detection of fetal body movements from multichannel magnetocardiographic signals. *Physiological Measurement*, 40(8):085005, 2019.

[215] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological Measurement*, 31(4):513, 2010.

[216] K. Schneider and Maternal Fetal Medicine Study Group. S1-guideline on the use of CTG during pregnancy and labor. *Geburtshilfe und Frauenheilkunde*, 74(08):721–732, 2014.

[217] U. Schneider, A. Fiedler, M. Liehr, C. Kähler, and E. Schleussner. Fetal heart rate variability in growth restricted fetuses. *Biomedizinische Technik*, 51(4):248–250, 2006.

[218] U. Schneider, E. Schleussner, A. Fiedler, S. Jaekel, M. Liehr, J. Haueisen, and D. Hoyer. Fetal heart rate variability reveals differential dynamics in the intrauterine development of the sympathetic and parasympathetic branches of the autonomic nervous system. *Physiological Measurement*, 30(2):215, 2009.

[219] S. Shakespeare, J. Crowe, B. Hayes-Gill, K. Bhogal, and D. James. The information content of doppler ultrasound signals from the fetal heart. *Medical and Biological Engineering and Computing*, 39(6):619–626, 2001.

[220] M. Shao, K. E. Barner, and M. H. Goodman. An interference cancellation algorithm for noninvasive extraction of transabdominal fetal electroencephalogram (TaFEEG). *IEEE Transactions on Biomedical Engineering*, 51(3):471–483, 2004.

[221] C. Signore, R. K. Freeman, and C. Y. Spong. Antenatal testing–a reevaluation: Executive summary of a Eunice Kennedy Shriver National Institute of Child Health and Human Development workshop. *Obstetrics and Gynecology*, 113(3):687, 2009.

[222] M. G. Signorini, N. Pini, A. Malovini, R. Bellazzi, and G. Magenes. Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Computer Methods and Programs in Biomedicine*, 185:105015, 2020.

[223] C. V. Smith, J. P. Phelan, R. H. Paul, and P. Broussard. Fetal acoustic stimulation testing: A retrospective experience with the fetal acoustic stimulation test. *American Journal of Obstetrics and Gynecology*, 153(5):567–568, 1985.

[224] C. V. Smith, J. P. Phelan, L. D. Platt, P. Broussard, and R. H. Paul. Fetal acoustic stimulation testing: II. A randomized clinical comparison with the

nonstress test. *American Journal of Obstetrics and Gynecology*, 155(1):131–134, 1986.

[225] V. Smith, A. Nair, R. Warty, J. A. Sursas, F. da Silva Costa, and E. M. Wallace. A systematic review on the utility of non-invasive electrophysiological assessment in evaluating for intra uterine growth restriction. *BMC Pregnancy and Childbirth*, 19(1):230, 2019.

[226] S. F. V. Sondaal, J. L. Browne, M. Amoakoh-Coleman, A. Borgstein, A. S. Miltenburg, M. Verwijs, and K. Klipstein-Grobusch. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low-and middle-income countries: a systematic review. *PloS One*, 11(5):e0154664, 2016.

[227] P. Soothill, R. Ajayi, S. Campbell, and K. Nicolaides. Prediction of morbidity in small and normally grown fetuses by fetal heart rate variability, biophysical profile score and umbilical artery Doppler studies. *BJOG: An International Journal of Obstetrics and Gynaecology*, 100(8):742–745, 1993.

[228] D. B. Springer, L. Tarassenko, and G. D. Clifford. Logistic regression-HSMM-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, 63(4):822–832, 2015.

[229] B. Sriram, M. A. Mencer, S. McKelvey, E. R. Siegel, S. Vairavan, J. D. Wilson, H. Preissl, H. Eswaran, and R. B. Govindan. Differences in the sleep states of IUGR and low-risk fetuses: An mcg study. *Early Human Development*, 89(10):815–819, 2013.

[230] K. A. Stewart, S. M. Navarro, S. Kambala, G. Tan, R. Poondla, S. Lederman, K. Barbour, and C. Lavy. Trends in ultrasound use in low and middle income countries: A systematic review. *International Journal of Maternal and Child Health and AIDS*, 9(1):103, 2020.

[231] U. A. Stock and J. P. Vacanti. Cardiovascular physiology during fetal development and implications for tissue engineering. *Tissue Engineering*, 7(1):1–7, 2001.

[232] J. F. Strasburger, B. Cheulkar, and R. T. Wakai. Magnetocardiography for fetal arrhythmias. *Heart Rhythm*, 5(7):1073–1076, 2008.

[233] L. Stroux. *A Perinatal Monitoring System for Low Resource Settings*. D.phil. dissertation, University of Oxford, Oxford, UK, 2016.

[234] L. Stroux. *A perinatal monitoring system for low-resource settings*. PhD thesis, University of Oxford, 2016.

[235] L. Stroux and G. D. Clifford. An affordable multisensor perinatal monitoring concept - the importance of signal quality indices for successful mhealth implementation. In *WHO Global Forum on Medical Devices 'Priority Medical Devices for Universal Health Coverage'*, Geneva, Switzerland, 2013. WHO.

[236] L. Stroux and G. D. Clifford. The importance of biomedical signal quality classification for successful mHealth implementation. In *2014 Tech4Dev International Conference UNESCO Chair in Technologies for Development: What is Essential?* EPFL, Lausanne, Switzerland, 2014.

[237] L. Stroux, B. Martinez, E. Coyote, N. King, R. Hall-Clifford, P. Rohloff, and G. D. Clifford. An mHealth monitoring system for traditional birth attendant-led antenatal risk assessment in rural Guatemala. *Journal of Medical Engineering and Technology*, 40(7-8):356–371, 2016.

[238] L. Stroux, C. W. Redman, A. Georgieva, S. J. Payne, and G. D. Clifford. Doppler-based fetal heart rate analysis markers for the detection of early intrauterine growth restriction. *Acta Obstetricia et Gynecologica Scandinavica*, 96(11):1322–1329, 2017.

[239] R. Tapia-Conyer, S. Lyford, R. Saucedo, M. Casale, H. Gallardo, K. Becerra, J. Mack, R. Mujica, D. Estrada, A. Sanchez, and R. Sabido. Improving perinatal care in the rural regions worldwide by wireless enabled antepartum fetal monitoring: A demonstration project. *International Journal of Telemedicine and Applications*, 2015:3, 2015.

[240] F. Tetschke, U. Schneider, E. Schleussner, O. W. Witte, and D. Hoyer. Assessment of fetal maturation age by heart rate variability measures using random forest methodology. *Computers in Biology and Medicine*, 70:157–162, 2016.

[241] T. Todros, C. Preve, C. Plazzotta, M. Biolcati, and P. Lombardo. Fetal heart rate tracings: Observers versus computer assessment. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 68:83–86, 1996.

[242] B. Trudinger, C. Cook, L. Jones, and W. Giles. A comparison of fetal heart rate monitoring and umbilical artery waveforms in the recognition of fetal compromise. *BJOG: An International Journal of Obstetrics and Gynaecology*, 93(2):171–175, 1986.

[243] UNICEF, World Health Organization, the World Bank Group, and the United Nations. Levels and trends in child mortality. Report 2019, 2019. Retrieved from `https://data.unicef.org/resources/levels-and-trends-in-child-mortality/` on Jan 23 2020.

[244] UNICEF, World Health Organization, World Bank, and UN DESA Population Division. Low birthweight. country, reginal and global estimates, 2018. Retrieved from `https://data.worldbank.org/indicator/SH.DYN.NMRT?most_recent_value_desc=true` on Jan 21 2020.

[245] UNICEF, WHO, World Bank, United Nations. Levels and trends in child mortality estimation, 2013.

[246] J. Unterscheider, S. Daly, M. P. Geary, M. M. Kennelly, F. M. McAuliffe, K. O'Donoghue, A. Hunter, J. J. Morrison, G. Burke, P. Dicker, and E. C. Tully. Optimizing the definition of intrauterine growth restriction: the multicenter prospective PORTO study. *American Journal of Obstetrics and Gynecology*, 208(4):290–e1, 2013.

[247] K. R. Uquillas, B. H. Grubbs, A. E. Prosper, R. H. Chmait, E. G. Grant, and D. K. Walker. Doppler us in the evaluation of fetal growth and perinatal health. *Radiographics*, 37(6):1831–1838, 2017.

[248] C. E. Valderrama, F. Marzbanrad, R. Hall-Clifford, P. Rohloff, and G. D. Clifford. A proxy for detecting IUGR based on gestational age estimation in a Guatemalan rural population. *Frontiers in Artificial Intelligence*, 3:56, 2020.

[249] C. E. Valderrama, F. Marzbanrad, M. Juarez, R. Hall-Clifford, P. Rohloff, and G. D. Clifford. Estimating birth weight from observed postnatal weights in a Guatemalan highland community. *Physiological Measurement*, 41(3):025008, 2020.

[250] C. E. Valderrama, F. Marzbanrad, L. Stroux, and G. D. Clifford. Template-based quality assessment of the Doppler ultrasound signal for fetal monitoring. *Frontiers in Physiology*, 8:511, 2017.

[251] C. E. Valderrama, F. Marzbanrad, L. Stroux, B. Martinez, R. Hall-Clifford, C. Liu, N. Katebi, P. Rohloff, and G. D. Clifford. Improving the quality of point of care diagnostics with real-time machine learning in low literacy LMIC settings. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 1–11, 2018.

[252] C. E. Valderrama, L. Stroux, N. Katebi, E. Paljug, R. Hall-Clifford, P. Rohloff, F. Marzbanrad, and G. D. Clifford. An open source autocorrelation-based

method for fetal heart rate estimation from one-dimensional Doppler ultrasound. *Physiological Measurement*, 40(2):025005, 2019.

[253] P. Várady, L. Wildt, Z. Benyó, and A. Hein. An advanced method in fetal phonocardiography. *Computational Methods Programs Biomedical*, 71(3):283–296, July 2003.

[254] C. L. Velayo, K. Funamoto, J. N. I. Silao, Y. Kimura, and K. Nicolaides. Evaluation of abdominal fetal electrocardiography in early intrauterine growth restriction. *Frontiers in Physiology*, 8:437, 2017.

[255] A. N. Vest, G. Da Poian, Q. Li, C. Liu, S. Nemati, A. J. Shah, and G. D. Clifford. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiological Measurement*, 39(10):105004, 2018.

[256] S. M. Vijgen, M. E. Westerhuis, B. C. Opmeer, G. H. Visser, K. G. Moons, M. M. Porath, G. S. Oei, H. P. Van Geijn, A. C. Bolte, C. Willekes, and J. Nijhuis. Cost-effectiveness of cardiotocography plus ST analysis of the fetal electrocardiogram compared with cardiotocography only. *Acta Obstetricia et Gynecologica Scandinavica*, 90(7):772–778, 2011.

[257] J. P. Vogel, J. P. Souza, and A. M. Gülmezoglu. Patterns and outcomes of induction of labour in Africa and Asia: a secondary analysis of the WHO global survey on maternal and neonatal health. *PloS one*, 8(6):e65612, 2013.

[258] R. T. Wakai. Assessment of fetal neurodevelopment via fetal magnetocardiography. *Experimental Neurology*, 190:65–71, 2004.

[259] R. T. Wakai. The atomic magnetometer: A new era in biomagnetism. In *AIP Conference Proceedings*, volume 1626, pages 46–54. American Institute of Physics, 2014.

[260] S. N. Wall, A. C. Lee, W. Carlo, R. Goldenberg, S. Niermeyer, G. L. Darmstadt, W. Keenan, Z. A. Bhutta, J. Perlman, and J. E. Lawn. Reducing intrapartum-related neonatal deaths in low-and middle-income countries—what works? *Seminars in Perinatology*, 34(6):395–407, 2010.

[261] U. Wallwitz, U. Schneider, S. Nowack, J. Feuker, S. Bauer, A. Rudolph, and D. Hoyer. Development of integrative autonomic nervous system function: an investigation based on time correlation in fetal heart rate patterns. *Journal of Perinatal Medicine*, 40(6):659–667, 2012.

[262] H. Wang, Z. A. Bhutta, M. M. Coates, M. Coggeshall, L. Dandona, K. Diallo, E. B. Franca, M. Fraser, N. Fullman, P. W. Gething, and S. Hay. Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: A systematic analysis for the global burden of disease study 2015. *The Lancet*, 388(10053):1725–1774, 2016.

[263] Y. Wang, W. Liu, X. Ma, J. Bailey, H. Zha, L. Song, and S.-T. Xia. Iterative learning with open-set noisy labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8688–8696, 2018.

[264] T. M. Wardlaw. *Low birthweight: country, regional and global estimates*. Unicef, 2004.

[265] P. A. Warrick, E. F. Hamilton, R. E. Kearney, and D. Precup. A machine learning approach to the detection of fetal hypoxia during labor and delivery. *AI Magazine*, 33(2):79, 2012.

[266] M. Whitworth, L. Bricker, and C. Mullan. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database of Systematic Reviews*, (7), 2015.

[267] A. Wolfberg. The future of fetal monitoring. *Reviews in Obstetrics and Gynecology*, 5(3/4):e132–6, 2012.

[268] World Health Organization. Perinatal mortality: a listing of available information, 1996. Retrieved from https://apps.who.int/iris/handle/10665/60977 on Jan 11 2020.

[269] World Health Organization. Global status report on noncommunicable diseases 2014, 2014. Retrieved from https://www.who.int/nmh/publications/ncd-status-report-2014/en/ on Jan 11 2020.

[270] World Health Organization. Levels and trends in child mortality 2014, 2014.

[271] World Health Organization. WHO compendium of innovative health technologies for low-resource settings, 2014. Retrieved from https://www.who.int/medical_devices/innovation/compendium/en/ on Jan 11 2020.

[272] World Health Organization. Maternal and perinatal health, 2016. Retrieved from http://www.who.int/maternal_child_adolescent/topics/maternal/maternal_perinatal/en/ on Feb 20 2017.

[273] World Health Organization. The neglected tragedy of stillbirths, 2016. Retrieved from https://www.who.int/reproductivehealth/topics/maternal_perinatal/stillbirth/Lancet-series/en/ on Jan 11 2020.

[274] World Health Organization. WHO recommendations on antenatal care for a positive pregnancy experience, 2016. Retrieved from https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/anc-positive-pregnancy-experience/en/ on Oct 11 2019.

[275] World Health Organization. WHO recommendation on continuous cardiotocography during labour, 2018. Retrieved from https://extranet.who.int/rhl/topics/preconception-pregnancy-childbirth-and-postpartum-care/care-during-childbirth/care-during-labour-1st-stage/who-recommend

ation-intermittent-fetal-heart-rate-auscultation-during-labour on May 4 2020.

[276] World Health Organization. WHO recommendation on intermittent fetal heart rate auscultation during labour, 2018. Retrieved from https://extranet.who.int/rhl/topics/preconception-pregnancy-childbirth-and-postpartum-care/care-during-childbirth/care-during-labour-1st-stage/who-recommendation-continuous-cardiotocography-during-labour on May 4 2020.

[277] World Health Organization, UNICEF, UNFPA, World Bank, and UN-DESA Population Division. Trends in maternal mortality: 1990 to 2013, 2014. Retrieved from https://www.who.int/reproductivehealth/publications/monitoring/maternal-mortality-2013/en/ on Nov 11 2017.

[278] J. Wrobel, J. Jezewski, and K. Horoba. Coping with limitation of bedside measurement instrumentation for reliable assessment of fetal heart rate variability. In *Information Technologies in Biomedicine*, pages 307–314. Springer, 2008.

[279] S. Wu, Y. Shen, Z. Zhou, L. Lin, Y. Zeng, and X. Gao. Research of fetal ECG extraction using wavelet analysis and adaptive filtering. *Computers in Biology and Medicine*, 43(10):1622–1627, 2013.

[280] J. Wyatt. Appropriate medical technology for perinatal care in low-resource countries. *Annals of Tropical Paediatrics*, 28(4):243–251, 2008.

[281] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, pages 802–810, 2015.

[282] M. Y. Yakoob, E. V. Menezes, T. Soomro, R. A. Haws, G. L. Darmstadt, and Z. A. Bhutta. Reducing stillbirths: behavioural and nutritional interventions before and during pregnancy. *BMC Pregnancy and Childbirth*, 9(1):S3, 2009.

[283] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.

[284] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland. The HTK book. Technical report, Cambridge University, Cambridge, UK, 2009.

[285] S.-Z. Yu. Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215–243, 2010.

[286] J. Zupan. Perinatal mortality in developing countries. *New England Journal of Medicine*, 352(20):2047–2048, 2005.