

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:

Camilo Ernesto Valderrama Cuadros

Date

A framework for gestational development tracking using 1D-Doppler ultrasound
signals in rural Guatemala

By

Camilo Ernesto Valderrama Cuadros
Doctor of Philosophy

Computer Science and Informatics

Gari D. Clifford, DPhil.
Advisor

Faezeh Marzbanrad, Ph.D.
Advisor

Qiao Li, Ph.D.
Committee Member

Babak Mahmoudi, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

A framework for gestational development tracking using 1D-Doppler ultrasound
signals in rural Guatemala

By

Camilo Ernesto Valderrama Cuadros
B.Sc., Universidad Icesi, Colombia, 2011
M.Sc., Universidad Icesi, Colombia, 2014
M.Sc., Emory University, GA, 2018

Advisor: Gari D. Clifford, DPhil.
Advisor: Faezeh Marzbanrad, 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
2020

Abstract

A framework for gestational development tracking using 1D-Doppler ultrasound signals in rural Guatemala

By Camilo Ernesto Valderrama Cuadros

Guatemala shoulders the burden of one of the highest perinatal mortality rates in Latin America, particularly among rural indigenous (Mayan) communities. This thesis aims to develop a fetal monitoring system for a highland community in Chimate-nalco, Guatemala, by designing methods to analyze one-dimension Doppler ultrasound signals (1D-DUS) recorded with a low-cost mobile health system piloted in the same community.

To that end, signal processing and machine learning techniques were used to address issues found in the pilot and initial randomized control trial of the mobile health system. In particular, four related pieces of research were undertaken. First, a signal quality method was developed to assess the utility of the 1D-DUS, achieving an F1-score higher than 90% to classify the signals into five distinct types of error. Then, an autocorrelation-based method to estimate fetal heart rate (FHR) from 1D-DUS was developed using a dataset simultaneously recorded with a fetal electrocardiogram. The method was shown to be generalizable, accurately estimating FHR for two independent datasets, including one collected in the Guatemalan highland community that is the focus of this study. Third, estimation of birth weight from observed post-natal weight was performed as a proxy to identify small-for-gestational-age births, achieving similar results to those provided by the Guatemalan government for the region of study. Fourth, fetal heart rate variability indices were combined with maternal blood pressure readings to estimate gestational age using a supervised support vector regression approach. The estimations resulted in a mean absolute estimation error of 0.8 months, which is comparable to previous works developed in industrialized environments while requiring only an inexpensive transducer and a self-inflating blood pressure device.

This thesis provides low-cost approaches for identifying high-quality 1D-DUS signals, estimating FHR, and in turn, estimating gestational age in order to identify potential cases of Low Birth Weight, Small for Gestational Age, or Intrauterine Growth Restriction. The work empowers traditional birth attendants with a decision support system to identify patients with possible pregnancy-related abnormalities requiring professional medical assistance.

A framework for gestational development tracking using 1D-Doppler ultrasound
signals in rural Guatemala

By

Camilo Ernesto Valderrama Cuadros
B.Sc., Universidad Icesi, Colombia, 2011
M.Sc., Universidad Icesi, Colombia, 2014
M.Sc., Emory University, GA, 2018

Advisor: Gari D. Clifford, DPhil.
Advisor: Faezeh Marzbanrad, 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
2020

Acknowledgments

I am extremely grateful to my advisor Professor Gari D. Clifford, for giving me the opportunity to work and learn from him throughout my doctoral studies. His helpful advice and his invaluable insight into machine learning applied to the medical field allow me to strive during my research project. I am also thank you for his patient and valuable help with editing my manuscripts. Finally, I thank Dr. Clifford for financial support during Spring 2020.

I would also like to extend my sincere thanks to my co-advisor, Dr. Marzbanradz. I really appreciate her patient, kindness, and immense contribution during our manuscripts writing process. I'm also deeply indebted to her generosity of covering the fee of my first publication in my doctoral studies.

I very much appreciate my research collaborators, whose help was crucial for the culmination of my research project. Dr. Lisa Stroux for her guidance in the first years, and research support in our common publications. Her doctoral thesis was an excellent starting point for all the methods developed in this thesis. Dr. Peter Rohloff, and the Waku Kawoq team, for guidance and valuable suggestions during the development of the projects presented in this thesis. Nasim Katebi for our research discussions and help in our common project. Chengyu Liu and Elianna Paljug for their help in the annotation process of the datasets.

Special thanks to my mother for all her advice and emotional support during these challenging years. Thanks to my sister, aunts, and cousins for all their support from Colombia. Thanks to my girlfriend, Maria, for her compression and encouragement. Thanks to my friend Andres for our mutual support in this research journey. To Alex, Samantha, Giulia, Erick, and Andoni for all the fun adventures in these years. To Pradyunma and Ayse for sharing with me all their wisdom during many lunches.

I very much appreciate the BMI staff, in particular, Robert and Jim, for their help with the technical issues, and Cathy for her administrative support. I am also

grateful to Grace Song, who kindly edited some passages of this thesis. I would like to express my deepest appreciation to my committee, Dr. Li and Dr. Mahmoudi, for their time and feedback as assessor of this project.

Finally, I would like to recognize the financial support that I received from the Colombian Administrative Department of Science, Technology, and Innovation (Colciencias) from 2015 to 2019 throughout the Fulbright scholarship. Likewise, I would like to thank the Laney Graduate School for the additional stipend received from 2015 to 2019 and the total support during Fall 2019.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aim of this thesis	3
1.3	Thesis outline	5
1.4	List of publication	6
2	Background	8
2.1	Abstract	8
2.2	Introduction	9
2.3	Common pregnancy complications	9
2.3.1	Intrauterine growth restriction	10
2.3.2	Preeclampsia and high blood pressure	10
2.3.3	Gestational diabetes	11
2.3.4	Infections	11
2.3.5	Preterm birth	12
2.3.6	Pregnancy loss/miscarriage	12
2.3.7	Stillbirths	12
2.4	Traditional physical exams	13
2.4.1	Fundal height	13
2.4.2	Fetal movement counting	13

2.4.3	Maternal blood pressure monitoring	14
2.5	Biochemical tests	14
2.5.1	Estriol assays	15
2.5.2	Human placental lactogen	15
2.6	Fetal cardiac assessment	15
2.6.1	Fetal cardiac circulation	15
2.6.2	Control of fetal heart rate	16
2.6.3	Fetal heart monitoring techniques	18
2.7	Imaging	25
2.7.1	Ultrasound imaging	25
2.7.2	Fetal echocardiography	30
2.7.3	Fetal Magnetic Resonance Imaging	32
2.8	Summary of fetal heart monitoring techniques	34
2.9	Perinatal mortality and fetal monitoring in low-and middle income countries and resource-constrained region	35
2.9.1	Mobile technology for perinatal care	39
2.10	Discussion and Conclusion	43
3	Signal quality method for assessing 1D-DUS in a clinical environ- ment	45
3.1	Abstract	45
3.2	Introduction	46
3.3	Methods	48
3.3.1	Database	48
3.3.2	Segment selection	49
3.3.3	Preprocessing	50
3.3.4	Template-based quality assessment of 1-D Doppler Ultrasound	51
3.3.5	Sample Entropy and Power Spectrum Density (PSD)	56

3.3.6	Feature Vectors	57
3.3.7	Classification	57
3.3.8	Method performance assessment	58
3.4	Results	60
3.4.1	Feature Selection	60
3.4.2	Test Set Performance	61
3.4.3	Performance of classifier on intermediate quality segments	61
3.5	Discussion	62
3.6	Conclusion	65
4	Data capture errors in the 1D-DUS recorded in the field	66
4.1	Abstract	66
4.2	Introduction	67
4.3	Methods	71
4.3.1	Data acquisition	71
4.3.2	Description of data	72
4.3.3	Preprocessing	72
4.3.4	Class annotation	73
4.3.5	Feature extraction	75
4.3.6	Feature selection	82
4.3.7	Classification	83
4.3.8	Performance assessment	83
4.3.9	Processing time	86
4.4	Results	87
4.4.1	Model performance	87
4.4.2	Feature selection and real time performance	89
4.5	Discussion	91
4.6	Conclusions	93

5	Fetal heart estimation from 1D-DUS signal	94
5.1	Abstract	94
5.2	Introduction	95
5.3	Methods	98
5.3.1	Databases	98
5.3.2	Manual Heart Rate Estimation	100
5.3.3	Datasets	105
5.3.4	Heart rate estimator	106
5.3.5	Performance assessment	113
5.4	Results	116
5.4.1	Optimized parameters and validation stage	116
5.4.2	Oxford dataset	117
5.4.3	Guatemala RCT dataset	118
5.4.4	Comparison between Oxford JR and Guatemala RCT datasets	120
5.5	Discussion	121
5.5.1	Interpretations of Findings	121
5.5.2	Study Limitations	123
5.5.3	Future Directions	125
5.6	Conclusion	125
6	Estimating birth weight from postnatal weights	126
6.1	Abstract	126
6.2	Introduction	127
6.3	Background	129
6.3.1	Infant weight development	129
6.3.2	Weight curve models	130
6.3.3	Birth weight in Guatemala	132
6.4	Methods	132

6.4.1	Database	132
6.4.2	Preprocessing	133
6.4.3	Fitting models	134
6.4.4	Comparison with other models	136
6.4.5	Estimating birth weight	136
6.4.6	Comparison of the estimated birth weight	137
6.4.7	Identification of low birth weight	137
6.5	Results	137
6.5.1	Preprocessed data	137
6.5.2	Fitted models	138
6.5.3	Comparison to other models	139
6.5.4	Estimation of birth weight and identification of low birth weight	139
6.6	Discussion	142
6.7	Conclusion	145

7 Estimating gestational age from 1D-DUS and maternal blood pressure recordings 146

7.1	Abstract	146
7.2	Introduction	147
7.3	Background	149
7.4	Methods	152
7.4.1	Databases	152
7.4.2	Deriving the FHR signal	155
7.4.3	Features used for gestational age estimation	157
7.4.4	Estimation of gestational age	163
7.4.5	Detecting possible cases of IUGR	167
7.5	Results	168
7.5.1	Training/Validation performance	168

7.5.2	Ranking the features	169
7.5.3	Testing performance	170
7.5.4	Comparing features for the estimation of GA	172
7.5.5	GA estimation errors as a function of birth weight	173
7.6	Discussion	174
7.6.1	Interpretations of Findings	174
7.6.2	Study Limitations	176
7.6.3	Future directions	178
7.7	Conclusion	178
8	Conclusion	179
8.1	Summary and contributions	179
8.2	Limitations	181
8.3	Future work	182
8.3.1	Online functionalities	182
8.3.2	Segmentation of 1D-Doppler ultrasound signals into beat-to-beat intervals	183
8.3.3	Deep learning for improving gestational age estimations and detection of IUGR cases	184
Appendix A	Table of all feature combinations for quality assessment	185
Appendix B	Algorithm for determining periodicity	189
Bibliography		190

List of Figures

3.1	Duration in minutes of the total number of records per subject. . . .	49
3.2	Number of poor and good quality 3.75 s segments for each of the subjects for which all three annotators agreed on labels.	51
3.3	Overview of the template-based quality method for the Doppler signal	52
3.4	Distributions of classifier probability outputs for DUS segments of test set for each of the four classes (n.u. stands for normalized units). The threshold of belonging to the Good class was fixed at 0.56 for the classifier. The majority of the distribution of the Good and Mostly Clean classes lies above this threshold, whereas the majority of Poor and Mostly Noise classes lies below this threshold, as was expected. The probability distributions were smoothed using a normal kernel function [53].	62
4.1	mHealth monitoring scenario. Providing traditional birth attendants (TBAs) with signal quality feedback at point of data collection is crucial for adequate referral, diagnosis and treatment.	70
4.2	Illustration of the Doppler-mobile phone recorder used for data acquisition. The smartphone is completely audio- and pictogram-driven for low literacy populations. It attaches to an ultrasound device for recording, which in turn attaches to a low cost speaker so that the user has immediate audio feedback of the fetal heartbeat.	71

4.3	Box plots (median \pm IQR) of number of recordings by TBAs in the second and third trimesters, and over both the latter trimesters. By design, no recordings were performed in the first trimesters.	73
4.4	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 [351], 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 <i>Good Quality</i> , <i>Poor Quality</i> , <i>Interference</i> , <i>Silent</i> , <i>Talking</i> , or <i>Unsure</i>	74
4.5	Power spectrum of the DUS segments. Electrical interference presents pronounced peaks around 217 and in its harmonics (i.e. 434 Hz, 651 Hz, 868 Hz, etc) which are associated to the radio protocol GSM-TDMA [21]. There is also a pronounced peak around 800 Hz, which is a frequency associated with audio distortion caused by amplifiers [4].	80
4.6	Durations of recordings taken from each subject (pregnant mother) in the data set. Subjects are grouped by the TBA who made the recordings ordered by the quantity of data recorded. For each TBA, subjects are ranked by total quantity of recording time.	84
4.7	Error bars (median \pm IQR) per-class recall as a function of window length over 100 iterations of balanced bootstrap resampling of the unbalanced data.	87
4.8	Error bars (median \pm IQR) per-class precision as a function of window length over the 100 iterations.	87

4.9	Median F1 scores (\pm IQR) of μ F1 (dashed blue line) and MF1 (dashed red line) for the different combinations of the mRMR-selected features. Black solid line represents the corresponding median computation time required for extracting the features from a DUS segment of 3.5 s. All results are for 100 bootstrap iterations.	90
4.10	Median μ F1 (dashed blue line) and MF1 (dashed red line) scores (\pm IQR) for increasing numbers of features in the classifier. Only 85 features were considered, with individual execution time of less than 100 ms, and only 17 features were selected using mRMR. Black solid line represents the corresponding median computation time required for extracting the features from a DUS segment of 3.5 s. All results represent averages over 100 bootstrap iterations.	91
5.1	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) [38]. 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).	101
5.2	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.	103

5.3	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.	105
5.4	Manual FHR (FHR_m) estimations for datasets: (a) Leipzig dataset; (b) Oxford dataset; (c) Guatemala RCT dataset. The solid (red) lines represent the probability distributions smoothed using a normal kernel function [53]. The Leipzig and Oxford datasets appear to have continuous distributions, while the Guatemala RCT contained a much more pronounced quantization.	107
5.5	Flow diagram of algorithm used for finding the peak in the ACF to determine the heart rate period. Pseudocode for this flow diagram is provided in Appendix B.	111
5.6	ACF for a 3.75 1D-DUS segment. Vertical dashed and dotted lines indicates the respective start and end points of the window within which a search is made for the peak related to the heart rate period. Peaks inside the window search are compared using algorithm showed in Figure 5.5 to find the right time location for the signal period.	112
5.7	Error (ϵ) (difference) between manual FHR (h_m^f) and automatic FHR (h_a^f) of the LHU validation set. Shaded area shows the 10% PPA bounds (error no greater than 10 percent of the h_m^f). N stands for the number of segments in the validation set, and in parenthesis is the percentage of segment satisfying the PPA. The red line is the robust least square fit of ϵ along the manual FHR estimations ($\epsilon = -0.571 + 0.04h_m^f; r^2 = 0.99$).	116
5.8	Bland-Altman plots for the LHU validation set. Mean (μ) and standard deviation (σ) of ϵ were 0.34 and 3.67 BPM, respectively. Limit of agreements ($\mu + 2\sigma$) was $[-7, 7.68]$ BPM. RSME was 3.68 BPM	117

5.9	Maternal-Fetal difference in heart rate, for the validation set. This quantity was defined as the difference between maternal heart rate (h_m^m) and automatic FHR (h_a^f) estimation. The shaded area represents bounds in which automatic HR estimation is considered maternal instead of fetal based on the PPA (h_a^f estimation within 10% of the recorded h_m^m). Only one estimation out of the 291 was within the shaded area, indicating a possible maternal recording (from maternal arteries).	117
5.10	Error (ϵ) in fetal heart rate estimation discriminated by signal quality for the Oxford dataset: (a) poor quality, (b) intermediate quality, (c) good quality, (d) excellent quality. ϵ was defined as the difference between manual FHR (h_m^f) and automatic FHR (h_a^f) estimations. Shaded area shows the PPA error bounds (error no greater than 10 percent of the h_m^f). N stands for the number of segments in that signal quality class, and in parenthesis is the percentage of segment within the tolerance error bounds. The red line is the robust least square fit of ϵ along the manual FHR estimations (r^2 was 0.995, 0.997, 0.997, and 1.0 for poor, intermediate, good, and excellent quality, respectively).	118
5.11	Bland-Altman plots for the fetal heart rate estimation discriminated by signal quality for the Oxford dataset: (a) poor quality, (b) intermediate quality, (c) good quality, (d) excellent quality. Mean and standard deviation, level of agreement, and RSME for each class are shown in Table 5.5.	119
5.12	Error (ϵ) measured as the difference between manual (h_m^f) and automatic FHR (h_a^f) estimations for Guatemala RCT dataset. The shaded area represents the PPA bound, defined as no greater than 10 percent of the h_m^f . S stands for the number of segments, and in parenthesis is the percentage of segment satisfying the PPA bounds. Red line is the robust least square of ϵ along the manual FHR estimations ($\epsilon = -38.63 + 0.28h_m^f; r^2 = 0.9962$).	119

5.13	Bland-Altman plots for the Guatemala RCT dataset. Mean (μ) and standard deviation (σ) of ϵ were -0.17 and 6.63 BPM, respectively. Limit of agreements ($\mu + 2\sigma$) was [-13.44, 13.10] BPM. RSME was 6.64 BPM	120
5.14	Box plot of the error (ϵ), defined as the difference between manual heart rate (h_m^f) and automatic FHR (h_a^f), for Guatemalan RCT dataset grouped by gestational age in months. N indicates the number of segments in the GA group and in parenthesis is the percentage of segments satisfying the PPA 10% error tolerance. FHR estimations did not statistically significantly differ among any possible pair of gestational months (Bonferroni corrected p-value of 0.005 (0.05/10); two-sided Wilcoxon rank sum test).	120
6.1	Expected infant weight curves as reported by [19], showing a maximum loss weight around the third to fifth days after birth. After this nadir, the weight steady increases during the first months of life.	129
6.2	Retained (open circle) and discarded (filled red circle) observed weights for (a) male and (b) female newborns.	138
6.3	Reeds2 models fitted for (a) male and (b) female newborns against models generated using coefficients reported by Kim and Pollit (1987)[186] in a Taiwanese population, Berkey (1992)[45] in a Boston hospital sample, and Simondon <i>et al.</i> (1992)[335] in a Congo rural community.	140
6.4	Reeds2 models fitted for (a) male and (b) female in observed weights after birth against WHO weight growth chart percentiles. The percentiles showed for the WHO weight dataset corresponds to the 5th, 25th, 50th, 75th, and 95th.	141

6.5	Male newborns' birth weights estimated using the fitted second-order Reeds model. (a) Histogram of estimated birth weights with the threshold for identifying low birth weight (vertical dotted line), corresponding to the 14.29 percentile, or 2.64 Kg. (b) Comparison of the distribution of estimated birth weights (red) with the measured birth weights in the 2015 Guatemalan national survey for male newborns of rural Chimatenalgo (blue dashed line).	142
6.6	Female newborns' birth weights estimated using the fitted second-order Reeds model. (a) Histogram of estimated birth weights with the threshold for identifying low birth weight (vertical dotted line), corresponding to the 16.29 percentile, or 2.57 Kg. (b) Comparison of the distribution of estimated birth weights (red) with the measured birth weights in the 2015 Guatemalan national survey for female newborns of rural Chimatenalgo (blue dashed line).	143
7.1	Variation of the FHR from 22-38 weeks during pregnancy. Note that the vertical axis has an arbitrary offset. Average FHR does not drop by such a large amount each week during pregnancy, but rather it drops on average by about 15 BPM from week 25 to week 40 [180]. Adapted from [386]	150
7.2	Median error of the 100 repetitions against GA provided by the LMP method for the normal weight newborns (blue circles) and the LBW newborns (red crosses). For newborns with more than one visit, a line connects the median error along with the GA.	171
7.3	Median and interquartile range of the estimated GA for each label of the LMP method for the normal weight newborns and the LBW newborns.	172

7.4 Error in GA from 100 repetitions (δ) as the difference between GA based on the LMP and the estimated GA. The median error of the 100 repetitions for each recording (or visit) is displayed (triangles for LBW newborns; and circles for normal birth weight newborns). Robust least square (RLS) fits are also shown. 174

List of Tables

2.1	Summary of available fetal heart monitoring methods. The cost column represents the medical equipment cost from low (\$) to extremely high (\$\$\$\$). Thickness of the lines indicate the evidence (accuracy) and color indicates the time required for training operators (green: low; blue: moderate; cyan: considerable; red: high; magenta: extreme)	35
3.1	Median classification performance of the 100 five-fold cross validation balanced with bootstrapping. IQR indicates inter-quartile range; ‡ indicates a significant improvement (Wilcoxon rank-sum test, $P < 0.05$) of a given feature combination compared to using a combination with one less feature.	61
3.2	Performance of the classifier averaged over 100 five-fold cross validation runs balanced with bootstrapping for the test set (with) using SQI2, PSD ratio, and sample entropy (SQI2, PSD, H_s) as features.	61
4.1	Number of segments per class for each window length after annotation, and adjudication.	75

4.2	Wavelet frequency decomposition. Level 1 and 2 contain valve activities, reported to be in the range 990-3300 Hz for the ultrasound transducer used here [255]. Level 3 corresponds to wall velocities, which are located at range of 257-429 Hz, given the Doppler transducer used in this work [327]. Level 4 included frequency that is partly associated with fetal body movement [405]	78
4.3	Confusion matrix of the classifier assessed on test data for window length of 3.75 s, and repeated 100 iterations. At each iteration, the data are sampled randomly with replacement, balancing the classes of the training set.	88
4.4	Median global metrics of classifier on DUS segments (assessed on test data), based on 100 iterations. Recall, precision, and F1 score were aggregated for all classes using both micro-average and macro-average approaches.	88
4.5	Cumulative performance when adding individual features selected by the mRMR algorithm ranked in order of execution time on a Intel(R) Xeon(R) CPU E5-2660 v2 @2.20GHz processor for a 3.75 segment of DUS with individual computation times. The final column indicates the performance of the classifier using the feature and all features above it in the table. (Note that due to rounding, the cumulative execution time is not exactly the sum of the individual times.)	89
4.6	Median performance obtained on the held out test data by the retrained classifier using the selected features by the mRMR algorithm for a window length of 3.75 s, based on 100 bootstrap iterations.	89

4.7	Optimize feature vector selected by the mRMR algorithm applied to those features whose median computation time for extraction from a 3.75 second segment is less than 100 ms. The features are ranked based on the individual execution time. τ is the cumulative time to compute the features in the corresponding row and all the features in the preceding rows. G, P, I, T, and S represent <i>Good Quality</i> , <i>Poor Quality</i> , <i>Interference</i> , <i>Talking</i> and <i>Silent</i> segments of data respectively.	90
5.1	Summary of final datasets used for developing fetal heart rate estimator. 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 remaining two dataset were used as independent test sets. For the Oxford dataset, the number of good and excellent quality segments are shown in parenthesis.	105
5.2	Expected minimum and maximum values for measured change in frequency (Hz) from over a gestational age ranging from 20 to 40 weeks. RV and LV stands for right and left ventricular, respectively.	109
5.3	Parameter intervals used in the Bayesian optimization process.	113
5.4	Best tuple of parameters found by the Bayesian optimization process.	116
5.5	Mean and standard deviation, level of agreement, and RSME of the Bland-Altman analysis for the Oxford dataset discriminated by signal quality class.	118
6.1	Detail of infant weight development reported by previous published studies.	130
6.2	Average demographics for the data used in this study. For each metric, the standard deviation and the number of samples available for that variable are shown in parenthesis.	133
6.3	Maximum weights for each month based on the WHO child growth chart [396].	134

6.4	Details of the fitted curves on the observed postnatal weights, grouped by gender. The window length in days and the samples of each fitted curve are provided. The final (9th) column shows the L1-norm between a vector composed of the parameters of each model, (γ) , and the target vector given by average of the metrics provided in Table 6.1, (τ) , which corresponds to a nadir of 3.5 days; a weight loss at the nadir of 7.8%; a weight recovery duration of 11.2 days; and a weight gain rate of 30.3 g/day	139
6.5	Coefficients for the fitted curves shown in Table 6.4. The final (10th) column shows the MSE between the fitted curve and the observed postnatal weights.	139
6.6	Descriptive statistics of the estimated birth weights and the 2015 Guatemalan national survey reported newborns of rural Chimaternalgo [254]. Last two columns show the p-value of: (i) a two-sample t-test; and (ii) a paired two-sided Wilcoxon signed rank test. Null hypothesis tests were whether the two weight distributions have equal mean and median, respectively.	141
7.1	Average demographics for the data used in this study. For each metric, the standard deviation and the number of patients available for that variable are shown in parenthesis.	155
7.2	Number of visits per gestational age (GA) taken with the last menstrual period (LMP) method.	155

7.3	Features used for GA estimation. The features included temporal features as basal heart rate; STV; LTV; II; LTV; STV/LTV ratio; ACC; DCA; the number of accelerations per minute; and the variance, mean, standard deviation, skewness, kurtosis, PNN5, root mean square difference of successive interbeat sequence. Features also included complexity features and spectral features as ApEn, fractal dimension, Lyapunov exponent, LF, MF, HF, LF/(MF+HF) ratio, Tone-Entropy, and Generalized mutual entropy. Finally, maternal hemodynamic formulas were also included.	158
7.4	Detail of maternal hemodynamic formulae calculated using the SBP, DBP and MHR taken with the self-inflating blood pressure device. . .	163
7.5	Number of visits per gestational age (GA) for the 104 normal birth weight training set, the 25 normal birth weight test set, and the 24 low birth weight test set.	164
7.6	Mean absolute errors (MAE) of the 50 trial five-fold cross validation for the Elastic Net, SVR, and Gradient boosting tree. For each model, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided.	168
7.7	Feature ranking obtained after averaging the individual 250 feature ranking resulted in the 50 trial five-fold cross-validation.	169
7.8	Mean absolute errors (MAE) of the 50 trial five-fold cross validation for the SVR using the top10, top15, and top20 features. For each model, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided.	170

7.9 Mean absolute errors (MAE) of the 100 trials on the test (held-out) normal birth weight and LBW newborns. For each type of newborns, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided. A † indicates a significant difference between the median GA estimations of the normal and LBW newborns for the 100 repetitions (two-sided Wilcoxon rank-sum test; $\alpha = 0.05$) 171

7.10 Median of the mean absolute errors (MAE) of the 100 trials on the test (held-out) normal birth weight and LBW newborns using separately the one-dimension Doppler ultrasound (1D-DUS) based features and maternal blood pressure (MBP) based features. The median MAE of each source of feature were compared to the MAE of combining both sources shown in Table 7.9. The comparison was performed using a one-sided Wilcoxon rank-sum test with the alternative hypothesis that the median MAE of the individual feature source is lower than the median MAE of combining features of both sources. * indicates a statistically significant difference regarding (left one-sided Wilcoxon rank-sum test; $\alpha = 0.05$) 173

A.1 Median classification performance of the 100 five-fold cross validation balanced with bootstrapping for all the possible feature combinations. The table is grouped for feature vectors of the same length. For each combination of features, the median and interquartile range of the accuracy rate of the 100 repetitions are shown. 185

List of Algorithms

Algorithm for determining periodicity of 1D-DUS signal	189
--	-----

Chapter 1

Introduction

This thesis contributes to approaches for fetal monitoring in low- and middle-income countries (LMICs) by developing methods to analyze one-dimension Doppler ultrasound (1D-DUS) signals recorded with a low-cost mobile health (mHealth) system, previously introduced in a rural highland community in Chimatenalgo, Guatemala [354]. These functionalities were defined based on issues identified during pilot research and a randomized control trial (RCT) in rural Guatemala [232, 231, 354, 351, 354].

1.1 Motivation

In LMICs, perinatal death is an alarming burden, accounting for around 98% of total perinatal deaths [408]. Among LMICs, Guatemala suffers one of the highest perinatal morbidity and mortality rates in Latin America, particularly affecting the Mayan population [367, 367]. This high burden is a result of barriers, such as economic status, language, and culture, that limit the access to professional medical care and routine perinatal screening. Due to these barriers, pregnant indigenous Guatemalan women are usually attended by lay midwives, or traditional birth attendants (TBAs), who lack the training to identify abnormal fetal development [177, 231].

One sign of abnormal development commonly associated with perinatal death is a low birth weight (LBW) (≤ 2.5 Kg) [205]. In LMICs, around 60% of LBW newborns are due to intrauterine growth restriction (IUGR), preventing the fetus from reaching its full growth potential [205], and associated with lower neurodevelopmental scores and other negative sequelae [258, 207]. IUGR accounts for 52% of stillbirths [340], and has a prevalence of between 9-11% in LMICs [89, 205].

In high-income countries, obstetricians detect suspicious cases of IUGR by Doppler sonography [213]. However, Doppler imaging requires specialized medical professionals and expensive equipment, thus limiting its use in LMICs [399].

One alternative method to detect IUGR is monitoring fetal heart rate (FHR), from which a nonstress test can be performed [289]. For example, a lack of accelerations indicates possible cases of IUGR [226]. In more detailed approaches, indexes derived from fetal heart rate variability (fHRV) analysis can be used to identify heart rate accelerations associated with normal fetal growth [15].

The potential of detecting IUGR cases with fHRV has been shown in recent works. Specifically, Stroux *et al.* [355] analyzed a total of 1,163 IUGR cases and 1,163 healthy controls using fHRV indexes derived via computerized cardiotocography (CTG). The comparison yielded significant differences between the two groups, achieving an area under the receiver operator curve (AUC) of 76% for predicting IUGR cases. Similarly, Signorini *et al.* [333] achieved an accuracy rate of 90% in discriminating between 60 IUGR and 60 healthy fetuses using a random forest trained on twelve time, spectral, and nonlinear indices extracted from CTG signals.

Due to the potential of fHRV analysis to identify possible IUGR cases, Stroux *et al.* [354] introduced an affordable project aimed to provide fetal monitoring to a Guatemalan rural community. This project introduced, in collaboration with a local non-governmental organization (*Wuqu' Kawoq* | *Maya Health Alliance*), an intuitive smartphone-based decision support system designed for low-literacy TBA. This

mHealth fetal monitoring system consists of a low-cost 1D Doppler device connected to a smartphone running a mobile application, and a standard oscillometric blood pressure monitor. The app guides the user to find the fetus and to record fetal cardiac activity by using the Doppler transducer. The app also guides the user through a check-list - the Pregnancy Risk Assessment Monitoring System (PRAMS) [64]. 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 or referral to an appropriate healthcare facility, if needed.

This mHealth technology has provided TBAs with decision support, and through cellular connectivity, has linked their traditional procedures with a formal healthcare referral process [232, 231, 354, 351]. Although the technology has so far proven to be feasible in a Guatemalan highland community, it needs some refinements to reduce noise and adapt the prediction algorithms to the Guatemalan population using the low-cost Doppler transducer [354, 232].

1.2 Aim of this thesis

This thesis aims to provide a method for identifying clean DUS data, estimating fHRV, and in turn, estimating gestational age in order to identify potential LBW, SGA, or IUGR. To that end, signal processing and machine learning techniques were used, with a focus on inexpensive computational techniques that can be transferred to the mHealth perinatal screening system introduced in [232, 231, 354, 351].

In order to achieve our final aim, the following novel research was performed:

- A method to ensure the quality of 1D-DUS at the point of care was developed.

During the first two release cycles of the app, around 40% of the recordings were of low quality. Low-quality signals can distort fetal well-being analysis [353, 220]. Therefore, the quality of recorded 1D-DUS signals should be assessed at the point of care so that users can either discard or retake poor quality recordings.

- A method for extracting fetal heart rate from 1D-DUS signals was then developed. Commercial CTG companies have not disclosed the source code used for extracting FHR from Doppler ultrasound. However, as FHR is needed to monitor fetal well-being, the second goal was to provide the mHealth system with a method to estimate FHR from 1D-DUS signals.
- A method for estimating birth weight from observed postnatal weights was then developed. This estimation was necessary as the first postpartum visit could happen either days or weeks after birth, thereby making the identification of possible LBW newborns problematic. By providing a method to estimate birth weight based on postnatal weight, cases of LBW may be identified. These LBW infants were then labeled as such to improve the training data for estimating gestational age.
- A method for estimating gestational age from 1D-DUS and maternal blood pressure recordings was then developed. Gestational age estimation is essential to identify growth restriction, assess fetal well-being, and determine the timing of delivery. By providing an estimation of gestational age from 1D-DUS and maternal blood pressure recordings, non-medically trained operators may be equipped with an objective metric to assess fetus gestational development.

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 in the field (see section 1.4).

Chapter 2 presents the background of this thesis. The chapter first presents techniques used for assessing fetal well-being, focusing on fetal heart monitoring. Then, the chapter describes the challenges LMICs face for providing fetal monitoring. The chapter concludes by presenting the potential of mobile health applications to overcome LMICs barriers and provide antenatal care.

Chapters 3 and 4 introduce approaches for assessing the quality of 1D-DUS signals recorded with a Doppler transducer. Chapter 3 presents the first approach, building on the work of Stroux et al. [353], to classify 1D-DUS signals as usable (good) or unusable (poor) quality. Chapter 4 extends this method to data collected in an LMIC by introducing an approach to classify 1D-DUS signals into five quality classes observed in rural Guatemalan communities during pilot testing.

Chapter 5 provides an autocorrelation-function (AC)-based method for estimating fetal heart rate from the 1D-DUS signals. FHR estimation is an essential step for assessing fetal well-being as it is crucial for calculating fHRV indices.

The remainder of the thesis is focused on estimating abnormalities in gestational growth. Chapter 6 explores the issue of assessing the birth weight to identify small-for-gestational-age (SGA) births. Since it is not always possible to record weight on the day of birth, and weight loss/gain in the following days is nonlinear, a regression-based approach to estimating birth weight from postnatal weight is described. The births identified as SGA were then excluded from the analysis in the final research chapter.

Chapter 7 then presents a supervised machine learning approach to estimate gestational age based on fHRV indexes extracted from 1D-DUS as well as metrics derived

from maternal blood pressure.

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., Clifford G. D. A review of fetal monitoring for well-being assessment, with a focus on LMICs. *Physiological Measurement*. Under review
(This publication appears in Chapter 2).
- Valderrama, C. E., Marzbanrad F., Stroux L., Clifford, G. D. (2017). Template-based quality assessment of the Doppler ultrasound signal for fetal monitoring, *Frontiers in Physiology*. 2017 Jul 18; 8, 1-10. doi: <https://doi.org/10.3389/fphys.2017.00511> [373]
(This publication appears in its entirety in Chapter 3)
- Valderrama, C. E., Marzbanrad F., Stroux L., Martinez B., Hall-Clifford R., Liu C., Katebi N. Rohloff P., Clifford G. D. (2018, June 20-22). Improving the Quality of Point of Care Diagnostics with Real-Time Machine Learning in Low Literacy LMIC Settings, Paper presented at the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18), San Jose-Menlo Park. doi: <https://doi.org/10.1145/3209811.3209815> [374]
(This publication appears in its entirety in Chapter 4).
- 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 26; 40(2), 025005.doi: <https://doi.org/10.1088/1361-6579/ab033d> [375]

(This publication appears in its entirety in Chapter 5).

- Valderrama C. E., Marzbanrad, F., Juarez M., Hall-Clifford R., Rohloff P., Clifford G. D. Estimating birth weight from observed postnatal weights in a Guatemalan highland community. *Physiological Measurement*. 2020 Mar 5; 41(3). doi: <https://doi.org/10.1088/1361-6579/ab7350> [372]

(This publication appears in its entirety in Chapter 6).

- Valderrama C. E., Marzbanrad, F., Hall-Clifford R., Rohloff P., Clifford G. D. Estimating gestational age from one-dimensional Doppler transducer and maternal blood pressure in a Guatemalan highland community. *Frontiers in Artificial Intelligence*. Under review.

(This publication appears in its entirety in Chapter 7).

Chapter 2

Background

2.1 Abstract

Little evidence exists for the utility of fetal monitoring during pregnancy, particularly during labor. Current practices in high-income countries are mostly determined by consensus ‘best practices’ from obstetric and gynecology professional societies. Protocols are often driven by the desire to be as safe as possible, at least at the moment, avoid litigation, and to detect any abnormalities or anomalies, regardless of the cost of downstream treatment. In high-resource settings, where there is no absence of overwhelming evidence to contradict standard practices, there may be an argument for this approach. In low-resource settings, action can be costly and lead to negative outcomes, both in the short and long term. It is, therefore, important to consider the evidence base and cost of fetal monitoring, particularly in the context of treatment and care in low-to-middle income counties.

This paper reviews the common methods for fetal monitoring, focusing on fetal cardiac assessment, which has shown to be a reliable indicator for assessing fetal well-being. It also presents the current situation of fetal monitoring in low-to-middle income counties, presenting the problems challenging the perinatal care access. Finally,

the paper presents how mobile technology can help to reduce barriers for accessing perinatal in poor-resource setting care.

2.2 Introduction

Complications during pregnancy and birth account for 40% of the worldwide perinatal and maternal deaths. In order to reduce the occurrence or severity of these fetal complications, obstetricians recommend fetal monitoring throughout pregnancy and at delivery [402, 213]. Antenatal fetal well-being assessment should be provided to both low-risk and high-risk pregnancies during perinatal care visits [226].

Fetal monitoring includes different techniques developed in the last four decades [213]. These techniques encompass biophysical methods that assess fetal growth and physiological function, as well as biochemical tests that assess endocrine functions of the placenta unit [226]. However, despite the widely use of fetal monitoring methods, there is still not enough evidence to prove the utility of them for improving perinatal outcomes [330].

2.3 Common pregnancy complications

Complications during pregnancy can affect maternal and fetal health. The most common pregnancy complications include, but are not limited to, the following: intrauterine growth restriction (IUGR) [205], as well as, maternal high blood pressure and preeclampsia, gestational diabetes, infections, preterm birth, pregnancy loss/miscarriage, and stillbirth [259].

2.3.1 Intrauterine growth restriction

IUGR is a pathological process inhibiting the fetus to achieve its genetic growth potential. IUGR complicates approximately 5-10% of the total pregnancies [122], and is associated with a high risk of postnatal mortality and a high morbidity rate, thereby affecting the infant's development [202].

Causes of IUGR include genetics factor, maternal malnutrition, fetal infection, uterine and placenta problems [260]. Among the causes, the most common is the capacity of the placenta to transport nutrients to the fetus, and permitting the release of fetal waste [260]. In detail, the placenta is expected to parallel grow with the fetus to allow for a normal vascularization [5]. However, when the placenta does not properly grow, placental function is compromised, thus affecting fetal-placenta flow.

IUGR can affect fetuses before, during, and after delivery. Its consequences include stillbirths, asphyxia or intrauterine hypoxia, meconium aspiration, and intrauterine fetal demise. Moreover, IUGR is linked to cognitive delay in childhood [24], and to cardiovascular diseases and diabetes in adulthood [34, 277].

2.3.2 Preeclampsia and high blood pressure

Preeclampsia is a two-stage pregnancy complication characterized by hypertension and proteinuria in the third trimester of pregnancy. The first stage of preeclampsia is abnormal placental perfusion resulting in placental insufficiency and the release of excessive amounts of placental materials into the maternal blood circulation. The first stage leads to the second stage, which is maternal hypertension and preeclampsia [150].

According to international standards [165], preeclampsia is defined as new-onset hypertension (systolic blood pressure >140 mmHg, or diastolic blood pressure >90 mmHg) and proteinuria on urine dipstick at least two occasions 4 hrs apart, or new-onset hypertension with/without proteinuria but with at least one other significant

symptom of end-organ dysfunction (e.g., visual changes, altered mental status, severe headache, severe abdominal pain, pulmonary edema, low platelets, elevated creatinine).

It has been reported that 10% of women have high blood pressure during pregnancy, and preeclampsia complicates 2% to 8% of pregnancies [101]. In low- and middle-income countries, results showed that 10% to 15% of direct maternal deaths are associated with preeclampsia [182, 101]. Essential factors in developing preeclampsia include genetic, immunologic, and environmental factors [325]. Hypertension complications are associated with increased rates of adverse maternal and fetal outcomes, including placental abruption, preterm delivery, fetal growth restriction, stillbirth, maternal death secondary to stroke and eclampsia, as well as future risk of hypertension, diabetes mellitus, and cardiovascular disease [17].

2.3.3 Gestational diabetes

Pregnant women can develop diabetes during gestation due to the hormonal changes. Thus, the body cannot properly digest glucose, thus accumulating it on the blood. Gestational diabetes has a prevalence around 13% worldwide [164].

Gestational diabetes can lead to preeclampsia, as well as increase the size of the fetus, which increases the risk for cesarean delivery [16].

2.3.4 Infections

Pregnant women are more likely to be affected for influenza, hepatitis E, herpes simplex, malaria, coccidioidomycosis, measles, smallpox, and varicella [190]. These infections need to be treated because they can cause serious consequences, such as: pregnancy loss/miscarriage and stillbirths, ectopic pregnancy, preterm labor and delivery, low birth weight, birth defects, and maternal health complications [135].

Sexually transmitted infections (STIs) also affect maternal and fetal health. Medi-

cal assistants should early test STIs on pregnant patients to provide timely treatment, and therefore, reduce STIs complications [57].

2.3.5 Preterm birth

Preterm labor is a delivery before the 37th week of gestation. Preterm labor is a risky event as fetal lung and brain functions are not completely developed until the 39-40 weeks [259].

Preterm birth can be a cause of infections, developing a shortened cervix, or previous preterm births [272]. The risk of developing a preterm birth can be reduced by taking progesterone, which has shown to reduce preterm births in around 30% of high risk pregnancies [245].

2.3.6 Pregnancy loss/miscarriage

Pregnancy miscarriage are fetal deaths occurring before the 20th week of gestation, however approximately 80% of miscarriages occur in the first twelve weeks [118]. The prevalence of miscarriage ranges from 10-25%, but the risk increases by maternal age [18]. The common signs associated with miscarriage are vaginal bleeding, or fluid or tissue passing from the vagina [118].

The main cause of miscarriage are chromosomal abnormalities [371]. Other causes of miscarriage includes maternal lifestyle (smoke, alcohol, obesity), maternal age, previous miscarriage, diabetes, and thyroid problems [118, 273].

2.3.7 Stillbirths

Fetal deaths after the 20th gestational week are stillbirths. Common causes of stillbirth are preeclampsia, poor fetal growth, placenta unit problems, umbilical problems, birth defects, infections, and maternal lifestyle [203].

Approximately half of the stillbirths occurred at intrapartum, particularly in developing countries [87]. At intrapartum, the most leading cause of death is asphyxia, which can be due to maternal preeclampsia, placental abruption, or umbilical cord accident [204, 387, 385, 133].

2.4 Traditional physical exams

2.4.1 Fundal height

Fundal height is one of the simplest methods. The distance between the upper border of the pubic symphysis and the uterine fundal is taken. The measurement is then compared with a reference value based on gestational age. This method has shown an adequate sensitivity and specificity for low birth weight fetuses; however, its use for fetal surveillance is limited [226].

2.4.2 Fetal movement counting

In this method, a pregnant mother registers the time at which she noticed more than ten movements [251]. Although this method is simple, it is not well recommended as it is prone to subjectivity[226], and is affected by several factors including movement type, anterior placenta, and nulliparity [249], thus challenging its capacity to prevent antepartum death. In fact, Neldam [262] reported that stillbirths did not statistically significantly differ between trained women and control groups. Furthermore, Mangesi *et al.*[227], reported that although fetal moving counting was able to identify fetuses at risk in five conducted studies, this identification did not translate into a significant reduction of stillbirths.

2.4.3 Maternal blood pressure monitoring

Hypertension disorders are major contributors to perinatal morbidity and mortality [257]. Hypertension is generally the earliest physical abnormality and the most important clinical symptom observed in preeclampsia. The diagnosis of gestational hypertension still relies on conventional clinic blood pressure measurements [28]. Therefore, blood pressure monitoring in pregnancy is one of the essential components of antenatal care, and the optimal and appropriate measure of blood pressure can not be overemphasized.

However, not all blood pressure devices have been adequately validated. In fact, Bello et al. reviewed 32 studies validating the accuracy of blood pressure of ambulatory, home, and clinic devices, founding that only 34% of the studies did not violate the medical validation protocols [40]. Thus, the authors concluded that medical professionals need to be aware of this limitation and use medical devices that have been validated for their target population.

2.5 Biochemical tests

Biochemical tests aim to measure the physiological changes of pregnancy, focusing on the woman's renal function, carbohydrate and protein metabolism, and hormonal patterns [363]. These methods are commonly not used as they required biochemical expertise and expensive lab equipment. Moreover, their sensitivity is not enough to detect the majority of pregnancy with an abnormal outcome [226]. As biophysical methods have less false-positive and false-negative than biochemical tests and are less costly, the use of biochemical tests has become redundant [226].

2.5.1 Estriol assays

The test is performed on a blood sample or in a 24-hour-period urine collection. The high variation in plasma estriol levels between time to time difficult the interpretation of this test [102]. A ratio below the 10th percentile (<14 g/mL) is associated with fetal distress in labor. However, evidence has not shown any benefit for using estriol assays, and therefore it is rarely used [226].

2.5.2 Human placental lactogen

The test measures the serum found in the placenta. During pregnancy, serum rises from 3 g/mL at 10-14 weeks to 10 g/mL at the 36th week. The serum acts as a human growth hormone, and therefore a low value indicates a possible risk[226]. However, human placental lactogen is not used for high-risk maternal surveillance.

2.6 Fetal cardiac assessment

Fetal cardiac activity is a reliable indicator of fetal well-being, being one of the most widely used methods [289]. By monitoring fetal heart rate, clinicians access useful information to identify non-reassuring fetal status, thereby allowing them to initiate intervention to avoid fetal distress or stillbirth [225].

2.6.1 Fetal cardiac circulation

Human cardiac development takes place within the first sixth gestational weeks [129], being functional by the eighth week [25]. The development starts with a primary heart tube, which evolves into the four-chamber heart structure composed of two atria and two ventricles.

Although the fetal heart has four chambers, fetal circulation is different from the adult circulation. This difference is mainly because, in the fetus, the oxygenated

blood comes from the placenta rather than the lungs[349].

In detail, the oxygenated blood coming from the umbilical vein is channeled to the liver and inferior vena cava via the ductus venosus. The majority of this well-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 [349, 121]. On the other hand, the remaining oxygenated blood passes from the right atrium to the right ventricular and subsequently is pumped to the pulmonary vein. As lungs are not functional in fetuses, a significant percentage of the blood in the pulmonary vein is passed to the aorta using the ductus arteriosus [349]. The blood in the aorta is channeled to oxygenate the fetal brain and tissues. Finally, deoxygenated blood is moved to the placenta via two umbilical arteries [121].

After birth, the foramen ovale closes. This occludes the ductus venosus and arteriosus, resulting in a separation of the pulmonary and circulatory functionalities [349].

2.6.2 Control of fetal heart rate

The fetal heart rate (FHR) represents the reciprocal of the interval between two successive fetal beats. Fetal beats are controlled by the sinoatrial (SA) node, the autonomic nervous system (ANS), and the baroreceptors and chemoreceptors.

The SA node, also called the pacemaker, is located in the right atrium. SA initiates the action potential, resulting in a contraction of the atriums during the beginning of the systolic phase [274]. Then, the action potential is propagated via the atrioventricular node (AV) to the bundle branches and the Purkinje fibers on the ventricular walls [274]. This impulse contracts the ventricles, pumping the blood to the pulmonary veins and the aorta, thus finishing the systolic phase [274]. Then, the impulse leaves the ventricles, starting the ventricular diastolic phase in which the ventricular walls are repolarized [274].

The electrical and mechanical events of a heart contraction generate the cardiac cycle, which is measured as the number of beats per minute. The pace of the cardiac activity is controlled throughout pregnancy by the ANS [121]. At fifteen weeks of gestation, the average FHR is 160 BPM. However, this rate decreases as the ANS matures. Specifically, the ANS is composed of the sympathetic and parasympathetic nervous systems. The sympathetic system accelerates the trigger rate of the SNA, thereby increasing the speed of the depolarization transmission to ventricles, and causing a faster FHR. The parasympathetic system, on the other hand, has the opposite capacity of reducing the FHR.

The balance between the sympathetic and parasympathetic systems sets the baseline of the heart rate. However, as the sympathetic system matures first than the parasympathetic one, the FHR is higher in the first months of gestation. As pregnancy advances, the parasympathetic system matures, and therefore the FHR becomes balance. The normal FHR during pregnancy is between 110 to 160 BPM [291].

The other two components controlling fetal heart rate are the baroreceptors and chemoreceptors. Baroreceptors are located in the aortic arch, carotid arteries, and brain stem. When blood pressure increases, baroreceptors stimulate the vagal nerve to slow the heart rate, thereby causing a reduction in the blood pressure. On the other hand, when blood pressure is decreased, baroreceptors reduces the parasympathetic tone to increase the fetal heart rate and blood pressure.

Similarly, the chemoreceptors, found in the aorta, carotid artery, and brain stem, also affect the fetal heart rate based on oxygen level sensing. Particularly, when the oxygen level decreases, the FHR is accelerated to increase the oxygen input rate from the placenta. However, when the reduction of the oxygen level is abrupt (hypoxemia), the chemoreceptors trigger a vagal response, resulting in a reduction of heart rate and an increment of the blood pressure.

2.6.3 Fetal heart monitoring techniques

Fetal heart monitoring can be performed by different technologies, which can be categorized into intermittent auscultation (IA) and electronic fetal monitoring (EFM) methods. IA techniques focus on verifying the fetal life or cardiac performance by counting the number of beats over short periods. The EFM, on the other hand, provides continuous information of the FHR for period longer than 10 minutes, but usually not longer than an hour. The aim of EFM is to provided indications of a fetal stress or distress based on FHR variability, commonly performed via cardiotocography (CTG)[247].

Blix *et al.*[51] performed an extensive revision in IA techniques, finding that the most common methods on this category are Pinard fetoscope, DeLee fetoscope, and hand-held Doppler devices. FHR estimations based on intermittent techniques differ from those of CTG because CTG machines use autocorrelation to frequently average FHR rather than counting beats during a specific period [233].

EFM techniques can be categorized into invasive and non-invasive methods. In the invasive mode, fetal electrocardiogram (fECG) is taken directly from the fetal scalp [32]. Although the invasive technique is more accurate than non-invasive ones, its use is limited to intrapartum when the membranes are ruptured. In contrast, non-invasive methods can be applied during the antenatal period. Non-invasive methods widely described in the literature are CTG [394, 30], abdominal fECG [317], phonocardiography (fPCG) [3], fetal magnetocardiography (fMCG) [123].

Fetal phonocardiogram

Fetal phonocardiogram (fPCG) is the electronic extension of Pinard and DeLee stethoscope. Similar to the stethoscope, fPCG is an IA technique in which a microphone is placed on the maternal abdomen to listen to fetal heart sound [317]. The audible heart sounds correspond to the closure of the fetal valves during the heart cycle

[191]. Thus, the closure of the mitral and tricuspid valves generate a sound called S1, whereas the closure of the semilunar valves (pulmonary and aorta) generates a sound called S2.

Both sounds, S1 and S2, have low acoustic energy and are affected by noise. The noise includes environmental noise, as well as other maternal and fetal physiological sounds, such as breathing, fetal moving, and maternal circulation [381].

fPCG can be used during the antepartum phase (gestational week ≥ 24) [381]. Although fPCG is an alternative to the traditional ultrasound used in perinatal control, it is not common in clinical practice as there are not enough commercial providers [143, 351].

Adithya *et al.*[3], in a revision of fPCG, showed that method is underutilized and has challenges related to the acquisition and processing of the signal. Nevertheless, the authors pinpointed the capacity of fPCG for extracting cardiac timing and intensity of fetal heart sounds, thereby providing useful information for diagnostic. More research needs to be conducted to improve fPCG before can compete with standard fetal monitoring methods, namely, CTG and ultrasound imaging.

One-dimension Doppler ultrasound

One-dimension Doppler ultrasound (1D-DUS) estimates FHR by measuring the Doppler shift between transmitted and received ultrasound beams due to the heart mechanical movement and blood flow. The Doppler magnitude frequency shift f_D , is described as [189]:

$$f_D = \frac{2f_o}{c}V \cos \theta, \quad (2.1)$$

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 angle the between ultrasound beam and the surface in radians.

In detail, the transmitted beam has to travel across various anatomical structures between maternal abdominal and the fetus's heart. The propagation starts with the maternal skin and subcutaneous tissue, then reaching the uterine muscles and amniotic sac to find the fetal heart finally [230]. The movement of the fetal heart reflects the ultrasound beam, thus propagating the ultrasound in the reverse order. The distance between the DUS transducer and fetal heart depends on the maternal phenotype, thus varying among nationalities [398], and socio-economic status and body mass index [252].

The shifted Doppler frequency is usually demodulated using phase-quadrature demodulation, in which the receiving signals is mixed with the carrier signals $\sin 2\pi f_o t$ and $\cos 2\pi f_o t$ [106]. The demodulated signal is used to estimated FHR by detecting its periodicity, which is considered as the time between heartbeats [327].

Doppler ultrasound includes two different modes: the continuous wave (CW) Doppler and the pulsed-wave (PW) Doppler. CW is found mainly in hand-held Doppler transducers, which provide an intermittent measurement to identify the fetal heart rate during a specific time. PW, on the other hand, is a continuous FHR recording used in standard CTG machines.

Doppler ultrasound can be used during intrapartum and antepartum after the 20th gestational week. It is an intermittent auscultation method to detect heart rate variability metrics to assess fetal well-being. Mahomed *et al.*[222], in a randomized control trial, showed that fetal monitoring using Doppler transducers can detect fetal distress. Previous studies have also shown that the use of Doppler ultrasound for fetal monitoring can identify a comparable number of fetal abnormalities than cardiotocography [155]. In fact, a randomized control study showed that intermittent auscultation methods, as one-dimension Doppler ultrasound or hand-held Doppler devices, are as safe as cardiotocography in low-risk pregnancies [92]. However, in high-income countries, Doppler transducers are used more as home-fetal monitoring

rather than a primary clinical method.

Cardiotocography

Cardiotocography (CTG) is the simultaneous recording of continuous fetal heart rate (FHR) and uterine contractions (UC). CTG is a standard method to provide fetal well-being assessment, aiming to reduce fetal mortality and morbidity and to estimate optimal time for delivery [140]. In order to record the FHR, the medical assistant applies a gel on the maternal abdomen and ultrasound transducer. The transducer is then moved across the maternal abdomen while listening to the Doppler signal [121].

CTG can be applied during antepartum starting at the 20th week of gestation [93]. However, CTG is most commonly provided after the 28th gestational week [322]. Although CTG is widely used, it has a high intra-and inter variability among interpreters [362], resulted in a low specificity. To reduce this subjectivity, CTG has been shifting from visual observation to computerized version to detect fetal abnormalities [84]. In a Cochrane revision [140], computerized CTG showed to significantly reduced perinatal deaths in comparison to traditional CTG. Furthermore, computerized CTG has been used in recent years to develop artificial intelligence methods to detect abnormal FHR patterns [70, 390], as well as critical diseases, such as IUGR [355, 333].

Although CTG is a common method for fetal monitoring in high-income countries, previous control trials have shown a lack of evidence for benefits associated with CTG, reporting an increment of 20% in cesarean interventions without improving fetal outcome [92]. In fact, CTG has not statistically significant improved perinatal outcomes in comparison to the use of traditional intermittent auscultation methods [179, 244]. Moreover, CTG machines are costly ($>$ \$ 450) [403] and require maintenance, supplies, and training for staff in their use and interpretation, thereby complicating CTG use in low-resource settings [402].

The FHR monitoring via CTG is usually performed by the nonstress test (NST). However, variations of the NST, such as the contraction stress test and acoustic stimulation, are also performed for FHR monitoring.

The nonstress test The nonstress test (NST) aims to find acceleration patterns on the FHR that are associated with fetus movement, and therefore to fetal well-being [226]. NST monitors the fetal heart rate patterns for at least 20 minutes without adding any stress for inducing any contraction. This test calculates the fetal baseline, which is later used to measure long-term and short-term variance, episodes of high and low variation, acceleration, and decelerations. A reactive output is found when more than two accelerations occurred within 20 minutes of observation, whereas a non-reactive one is when at most one acceleration occurred within a 40-minute period [15]. The NST has a low false-negative (0.3%), but a high false-positive (%50) rate [103].

Contraction stress test The contraction stress test (CST) is based on the premise that contractions induce a hypoxic state [330]. A healthy fetus could tolerate this hypoxic state, whereas a hypoxic fetus will demonstrate decelerations on his FHR [226].

During the test, contractions are induced using oxytocin or nipple stimulation [226]. Although this method has a low false-negative rate (0.04%) [120] and lower false positive rate than NST (30% vs. 55%) [15], CST has some disadvantages. These disadvantages include the need for an intravenous intervention, the risk of causing fetal hypoxia, and the induction of preterm birth [226].

Acoustic stimulation Acoustic stimulation is a substitution of the classical NST. Specifically, before starting the NST, the fetus is stimulated with an artificial larynx over the head for 3 seconds [226]. The artificial larynx produces a vibratory stimulus

of 80 Hz that makes a healthy fetus to move, thereby generating an acceleration in the FHR. The advantage of the acoustic stimulation is that the NST timing test can be decreased to 10 minutes [72], as well as reduce the number of non-reactive stated provided by regular NST without affecting readability [339, 338].

When a non-reactive state is found, the acoustic test is repeated in five minutes. If the test is still non-reactive, a fetal biophysical profile or CST is performed [226].

Fetal Electrocardiogram

Fetal electrocardiogram (fECG) measures the electrical activity of the fetal heart rate, recording the complex electrical signal of the heart. The ECG signal is composed of five components: P, Q, R, S, and T [274]. The P wave represents the depolarization of the atria and is followed by the atrial contraction (atrial systole). The atrial contraction is extended until the QRS complex, in which the atria are relaxed and repolarized. The complex QRS corresponds to the depolarization of the ventricles, in which ventricles are contracted when the peak R is reached. The QRS peak is followed by ventricle relaxation until the T wave. Finally, the T wave corresponds to the repolarization of the ventricles. fECG can be provided in an invasive manner at intrapartum or in a non-invasive manner since antepartum.

Morphology and beat-to-beat heart rate variability estimated from fECG have been shown to provide biomarkers for preeclampsia and growth restriction. Lakhno conducted a study on the effect of pre-eclampsia on FHRV recorded from 106 patients (30 healthy, 44 mild pre-eclampsia and 32 severe pre-eclampsia subjects) at 34-40 weeks of gestation. In this study, the modulated FHR variability indices captured the suppression of fetal biophysical activity and the development of fetal distress in severe preeclampsia [196]. The same author showed that preeclampsia cases exhibited shortening of fetal PQ and QT and increased T/QRS ratio was observed [195]. Also, in order to assess the effect of IUGR on parameters extracted from abdominal fECG

Velayo et al performed the study on 20 controls and 15 IUGR singleton pregnancies [382]. Clear P-QRST complexes were recognized in all cases and they demonstrated that in the IUGR fetuses, the QT intervals were significantly prolonged.

Invasive fetal Electrocardiogram Invasive fetal electrocardiogram (invasive-fECG) requires a rupture of the membrane to introduce electrodes via the cervix and place them on fetal scalp [152, 153, 154]. This technique filters the recorded signal, allowing visualizing the P and T waves, as well as the QRS complex.

As fetal surveillance, scalp fECG has been used as a complementary technique during intrapartum FHR monitoring [14, 269]. To this aim, the morphology of the ST waveform is analyzed (STAN) to find patterns associated with uterine complications [200, 199]. The use of STAN has shown to be effective for reducing neonatal encephalopathy [268, 269]. STAN can be provided since the 36th gestational week and is provided to high-risk pregnancies when a non-reactive CTG is obtained, or labor is induced by oxytocin.

Non-invasive fetal Electrocardiogram Non-invasive fECG measures the electrical fetal heart signal by placing electrodes on the maternal abdomen [194, 317]. Abdominal fECG can be provided since the 18th gestational week and can be provided both in antepartum and intrapartum.

The main purpose of the extracted fECG is to identify R peaks, and therefore perform beat-to-beat variability analysis [317]. Although abdominal fECG has low amplitude, abdominal fECG is able to estimate comparable or higher accurate fetal heart rate variability indexes than CTG, thus providing suitable FHR baseline, short- and long-term variations, and accelerations and decelerations [176, 76, 174].

Fetal Magnetocardiography

Fetal magnetocardiography (fMCG) uses a sensible superconducting sensor to measure the magnetic field of the fetal heart rate activity [289, 317]. fMCG obtained a comparable waveform to that of fECG. In fact, fMCG PQRST waveform is considered of high quality, and can be used to classify arrhythmia and detect congenital heart diseases [289].

fMCG can be used from the 20th gestational week. The procedure extraction is short and the patient has to be in a quiet state. However, fMCG is not routinely provided in perinatal care because it is expensive, needs a shielded room for measurement, and required highly skilled personnel [289]. On the other hand, alternative methods, such as the abdominal fECG or the hand-held Doppler devices, can be taken at any time during pregnancy, even at home by the patients themselves [289].

2.7 Imaging

Imaging has become a crucial tool in obstetrics since the last thirty years, particularly using ultrasound imaging [114], including two, three and four-dimensional Ultrasound [91].

2.7.1 Ultrasound imaging

Ultrasound imaging is the gold standard in fetal monitoring in high-income countries [403, 213]. It is used to evaluate fetal growth, fetal cardiac structure and function, as well as the blood circulation in the fetus, uterus, and placenta.

Ultrasound is usually provided during the second trimester, after the 20th gestational week. WHO, based on a Cochrane revision [392], recommends an ultrasound scan at early pregnancy (≤ 24 weeks) [403, 402]. Ultrasound imaging has shown to be effective for assessing the viability of the pregnancy, estimating gestational age,

detecting multiple pregnancies, and checking placental position [392]. However, there is not a strong evidence of the benefits of ultrasound scan on reducing perinatal mortality [100, 261]. Nevertheless, ultrasound imaging contributes to validate suspicious diagnosis without invasive and dangerous interventions, reduce labor induction for post-term pregnancy, and detect fetal malformation [392]. Moreover, ultrasound has shown potential in proper change of clinical management in at least 30% of cases [141].

There exist multiple parameters derivable from ultrasound scan of uteroplacental flow to identify pregnancies at risk. Previous studies demonstrated the impact of parameters related to placental perfusion, including resistance index, pulsatility index, or systolic/diastolic ratio from uterine artery [59, 305], fetal heart rate responses, and uteroplacental flows [157].

There is not total agreement regarding how often ultrasound scan should be provided throughout pregnancy. Some obstetricians recommend at least four ultrasound scans for normal pregnancies, whereas others recommend just one before the 24th gestational week [278]. When four visits are provided, the first one is given between the 10-14 to validate the pregnancy and estimate gestational age. Then, during the 18-22 week, anomaly scan and gestation age confirmation are provided. At the third scan, during 30-34 weeks, fetal growth is assessed, and finally, the last scan between 36 and 38 weeks focuses on the fetal weight and the position, helping to determine the optimal delivery mode. Ultrasound imaging equipment costs around \$ 10,000, and requires additional expenses for maintenance, supplies, replacement batteries, and staff training [399]. Given these high expenses, ultrasound scan has limited use in low-income countries [402].

The most common method for antenatal monitoring based on ultrasound imaging are Doppler velocimetry, Amniotic fluid volume, or Fetal biophysical profile.

Fetal biometry

Ultrasound imaging allows obstetricians to measure different fetal organs to estimate gestational age and fetal weight. To this aim, the most common measures are Biparietal Diameter (BDP), Femur Length (FL), Head Circumference (HC), Crown Rump Length (CRL), and Abdomen Circumference (AC) [226]. Using a combination of these measurements, it is expected to estimate fetal weight within 5% of the actual weight in 50% of the time, and within 10% in 80% of the time [122]. Fetal biometry measurements, however, are shown to be more effective during the first trimester. During the last two trimesters, fetal measurements are affected by genetic and nutritional factors, thus challenging the comparison with the reference values [298].

Fetal biometry measurements can differentiate between fetuses that are IUGR and those that are constitutionally small (SGA) [345]. Specifically, when an estimated weight is below the 10th percentile for gestational age, the fetus is considered as a growth-restricted based on the guidelines of the American College of Obstetricians and Gynecologists (ACOG) [370]. However, there is not a total agreement in the low-estimated weight percentile. Indeed, 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, thus concluding that the threshold should be below the 3rd percentile [369]. Furthermore, fetal biometry measurements ignore the fact that 10% of normal population are genetically predisposed to be small, thus increasing the false positive rate [122]. Hence, fetal biometry should be combined with methods assessing the fetal ANS physiology to increase the accuracy in detecting IUGR cases [122]. When IUGR is detected, the pregnancy is categorized as high-risk because this disease has long-term consequences. Therefore, in high-risk pregnancies, fetal surveillance is conducted to timing the delivery and reduced adverse perinatal outcomes.

Doppler velocimetry

Doppler velocimetry is used to evaluate pregnancies at risk of fetal compromise [364], such as growth restriction [8] or cardiovascular abnormalities [47]. This method assesses the blood flow in the umbilical arteries and vein to evaluate fetal reactions to placental insufficiency [336]. Healthy pregnancies rely on the placenta and fetus circulation as this is the way for transferring oxygen and nutrients, and for eliminating fetal waste products [226]. Healthy fetuses have a forward flow in the umbilical arteries during the cardiac cycle. In contrast, a fetus with increasing vascular resistance has low diastolic flow.

To assess umbilical flow, different indexes are determined, such as systolic and diastolic ratio, pulsatility index, and resistance index [15]. Higher indexes indicate a significant vascular resistance, thus implying that fetal health is on risk [47, 330].

The resistance indexes are mainly measured on the umbilical artery (UA), the middle cerebral artery (MCA), and the ductus venosus (DV) [250]. From these three areas, the UA Doppler is the only having randomized control trials that support its feasibility for fetal surveillance in high-risk pregnancies [9].

The UA Doppler measures the resistance in fetoplacental circulation flow, providing a pulsatility index (PI). In a healthy fetus, UA has a forward flow. However, an increment in the placental resistance obliterates the muscular arteries in the placental villi, resulting in a reduced diastolic flow [47]. This reduction causes an absent, and finally, a reverse of the fetoplacental circulation flow. Both absent and the reverse end-diastolic flows are visualized in the Doppler images. In the case of the absent end-diastolic flow (AEDF), the pronounced systolic peak is followed by an interruption, whereas in the reversed end-diastolic flow (REDF), the systolic peak is followed by a negative peak that indicates blood flow moving in the opposite direction. In fetal growth-restricted pregnancies with AEDF or REDF, the delivery is recommended at the 32nd gestational week [311].

In a fetus with restricted growth, an UA PI greater than the 95th percentile indicates a perinatal adverse outcome in previous randomized control trials [369, 275]. Likewise, the use of UA Doppler has also shown to be effective in reducing perinatal deaths, as well as to reduce induced deliveries [9].

MCA flow can detect problems caused by fetal hypoxemia in IUGR. A hypoxic state supplies the majority of oxygenated blood to the brain, heart, adrenal glands, thus affecting the peripheral circulation [370]. 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 ratio MCA PI/UA PI can detect adverse perinatal outcome [229], which are related to an increment of the diastolic flow due to hypoxia [253].

DV flow can detect a cardiac failure in IUGR, particularly in cases of early-onset fetal growth restriction [35]. It is a reliable marker to identify acidemia and stillbirth [35], caused by absent or reversed end-diastolic pressure at the ductus venosus. Although DV flow measurement has a moderate accuracy in detecting fetal compromise, previous works suggest that DV Doppler is not able to provide fetal surveillance itself [311]. Furthermore, DV Doppler does not provide any additional benefit than the use of traditional CTG for fetal monitoring [206]. Nevertheless, delaying delivery until finding an abnormality using DV flow could prevent neurological impairment in the long-term [124]. However, randomized control trials are needed to assess the benefits of DV flow measurement more accurately.

Other areas useful to manage fetal growth-restricted pregnancies are the blood flow measurements of the uterine artery, the aortic isthmus, umbilical vein, and the atrioventricular valves [250]. The uterine artery flow has shown the capacity to identify pre-eclampsia, and small-for-gestational-age neonates in high-risk pregnancies [311]. Aortic isthmus measures the balance between the impedance of the brain and systemic circulation, indicating a cardiac dysfunction when there is an abnor-

mal balance[81]. Umbilical vein measures fetal venous circulation, and a high value indicates an increased venous pressure that results in right-sided heart failure and myocardial hypoxia [266].

2.7.2 Fetal echocardiography

Fetal echocardiography is a noninvasive technique to assess fetal cardiac by using ultrasonography to examine fetal cardiac anatomy and function [132]. The primary use of fetal echocardiography is for detecting congenital heart disease (CHD), which is the most common abnormality in fetuses, with a prevalence of around 8 to 9 per 1,000 live births [151].

The accuracy and speed of fetal cardiac assessment have improved in the last decades by using advance techniques such as color Doppler [94]. During the assessment, obstetricians examine the heart structure, as well as direction, pattern, and velocity of the flow, and the volume flow [10]. The basic visualization of the chambers can be extended to include also the blood flow circulation among the chambers, in a technique called ‘five chambers views’. This extension has allowed increasing the sensitivity for detecting CHD cases in 5%, achieving a final sensitivity rate of 65% [11].

Fetal echocardiography also provides pulse wave Doppler, which is recommended for a complete evaluation of the fetal heart. The pulse wave shows the blood flow among atrioventricular, mitral, and tricuspid valves [2]. These valves have a dual-peak Doppler waveform: E-wave that is the passive diastolic filling, and the A-wave that is active diastolic filling (“atrial kick”) [2]. Healthy fetuses have A-waves greater than E-waves; however, E-waves amplitude increases throughout gestation. A higher increase of the E-wave/A-wave ratio is a sign of IUGR or congenital cystic adenomatoid malformation, causing, in some cases, mitral or tricuspid regurgitation [229, 224].

Modern techniques for echocardiography include three dimensional (3D) and four-

dimensional (4D) fetal heart assessment [91]. These techniques allow for real-time examination of the heart rate function, allowing a more accurate assessment of the heart structures to identify risk for cardiac anomalies [42, 68, 167].

Although fetal echocardiography is considered one of the most relevant techniques for fetal cardiac assessment, fetal echocardiography is an expensive method requiring qualified specialists to perform examinations [62]. Therefore, fetal echocardiography is only provided when maternal and fetal conditions indicate the need for it.

Amniotic fluid volume

The technique aims to measure the amniotic fluid, which is an indicator of the fetal physiological state. This fluid increases throughout pregnancy, passing from 60 mL to 900 mL. A decrease of amniotic fluid volume can be a cause of hypoxemia [330].

The most common method for estimating amniotic fluid volume is the 4-quadrant amniotic fluid index (AFI) [290, 312]. This method is measured using 2D-ultrasound as follows:

- Divide uterus in four quadrants using the umbilicus and linea negra as reference.
- Place the ultrasound transducer at each quadrant.
- Measure at each quadrant the vertical dimension of amniotic fluid pocket.

The summation of all the quadrant is defined as the amniotic fluid index and is compared with the Phelon's criteria [312], as:

- 5 cm or less: oligohydramnios.
- 5-8 cm: bordeline AFI (low).
- 8-24 cm: normal AFI.
- ≥ 25 cm: polyhydramnios.

Oligohydramnios is associated with fetal hypoxia and may lead to preterm birth, growth restriction, or stillbirth. Polyhydramnios, on the other hand, is associated with gastrointestinal obstruction, maternal diabetes, Rh incompatibility, or syphilis. The AFI, however, has low sensitivity for predicting both oligohydramnios and polyhydramnios, namely 10% and 29%, respectively [219].

Fetal biophysical profile

The fetal biophysical profile (BPP) combines five different biophysical variables into a single score. This score takes into account:

- Fetal breathing movement.
- Fetal movement.
- Fetal tone.
- Fetal reactivity.
- Fluid amniotic volume.

Each of these variables receives a value of at most 2 points, having a total maximum score of 10. A total score higher than 8 indicates normality in the fetus, 6 indicates an equivocal state, and lower than 4 indicates an abnormality in the fetus [15, 213]. As BPP combines different independent methods, its specificity and sensitivity for detecting abnormal fetus are higher than those of NST [226].

2.7.3 Fetal Magnetic Resonance Imaging

Fetal Magnetic Resonance Imaging (fMRI) aims to assess fetal abnormalities when a valid medical reason is found after sonographic screening [131, 119, 56, 208, 294].

The fMRI measures the magnetic moments of hydrogen nuclei in cells and tissue. Different components of the tissues translate into a different level of contrast in the

fMRI images, being water and fat the most contrast components [33]. The fMRI images are classified as weighted images (T1) and diffusion-weighted images (T2). In T1, fat has the highest contrast. On the other hand, water has the highest contrast in T2. The facilitation or restriction water movements describe the behavior of some pathological disorders that can be identified with T1 and T2 images.

The fMRI is taken in a supine position and has a duration between 10 to 35 minutes. The exam has to be guided by a radiologist and can be taken after the 18th week of gestation [162, 90].

As a complementary exam for sonography, fMRI provides useful information for the diagnosis, timing delivery, and counseling [26, 77, 63]. The primary indications of fMRI include:

- Central nervous system (CSN) is the most widely studied by fMRI. This includes assessment of congenital anomalies of the brain or skull [344, 343, 301], vascular brain problems [41, 58, 88], and spine abnormalities[130].
- Thoracic abnormalities. The fetal lung is filled with fluid and is easily observable in diffusion-weighted images. Therefore, by using fMRI any lung malfunction can be detected [161, 242].
- Gastrointestinal tract. Meconium can be visualized using either T1 or T2 fMRI images. The fMRI images allow examining the intestinal anatomy, measuring the level of meconium and amniotic fluid [314]. This measurement helps to detect bowel pathology and the level of obstruction.
- Genitourinary system. T2 fMRI images are relevant for visualization renal parenchyma and excretion system. fMRI has comparable results to detect renal abnormalities than ultrasound technology; however, fMRI is the only option when oligohydramnios or anhydramnios renders ultrasound is unreliable [293, 63].

- Fetal surgery assessment. fMRI can assist the planning of surgical interventions provided to the fetus during pregnancy [137]. Moreover, fMRI can be used to assess the fetus's status before and after the surgery.

Although fMRI has shown potential for fetal well-being assessment, it is still not commonly used for several reasons [60]. Firstly, sonography is satisfactory for routinely fetal screenings, thus relegating the use of fMRI. Secondly, obstetricians can not take the fMRI directly as it needs to be taken by a radiologist, thereby creating a personal logistic barrier. Finally, fMRI requires specialized medical equipment, making this technique more costly than CTG or sonography.

2.8 Summary of fetal heart monitoring techniques

Table 2.1 shows a summary of the fetal heart monitoring techniques. The table is sorted from the most affordable medical equipment to the most expensive one. For each technique, a timeline displays the gestational weeks in which it can be provided. The color and thickness of the timeline represent the needed operator skills and the evidence, respectively. The code color for required training time is: low (green), moderate (blue), considerable (cyan), high (red), and extreme (magenta). For the thickness, a wider line indicates more evidence for using the method as fetal monitoring.

Table 2.1: Summary of available fetal heart monitoring methods. The cost column represents the medical equipment cost from low (\$) to extremely high (\$\$\$\$\$). Thickness of the lines indicate the evidence (accuracy) and color indicates the time required for training operators (green: low; blue: moderate; cyan: considerable; red: high; magenta: extreme)

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

* GA \geq 24 weeks [381].

† GA \geq 20 weeks [289].

‡ GA \geq 20 weeks [289].

§ GA \geq 20 weeks [140].

¶ GA \geq 18 weeks [317].

|| Intrapartum (GA \geq 36 weeks) [269].

** GA \geq 20 weeks [402].

†† GA \geq 20 weeks [289].

‡‡ GA \geq 18 weeks [162].

2.9 Perinatal mortality and fetal monitoring in low- and middle income countries and resource-constrained region

Low-and middle-income countries (LMICs) account for around the 90% of total births, and 98% of the total perinatal deaths [50, 389, 395, 320]. Perinatal mortality rate, defined as the number of stillbirths and deaths in the first seven days of life per 1000 births, was 19 in 2018 in LMICs, being Sub-Saharan African and South-Asian

countries the most affected, with a rate of 28 and 26, respectively [368]. These figures could be even higher since, in LMICs, the registration of the number of deaths and births, as well as causes of deaths, is precarious [281, 215].

The alarming perinatal death rates were similar in high-income countries (HIC) at the beginning of the twentieth century. HICs reduced stillbirth and perinatal death rates by expanding antenatal care coverage, extending indication for cesarean sections, and introducing perinatal screening technologies (CTG, ultrasound, amnioscopy, amniocentesis, and pH-metry) [100, 136, 204, 117]. Moreover, the perinatal care services were organized in a regional network with neonatal intensive care units [283]. This network categorizes hospitals into three levels, in which pregnant women start from the basic services, and are referred to a higher category hospital when needed. However, due to the constraint of resources, LMICs have had a slow progress in implementing similar medical practices [283, 411].

In LMICs, perinatal deaths are mainly consequence of the lack of antenatal care visits. Experts recommend at least four perinatal visits to monitor the pregnancy accurately [402]. However, in LMICs, for various reasons, pregnant women do not attend prenatal visits. Finlayson and Downe [113] found that pregnant women are affected by both religion/cultural and economic aspects to attend perinatal visits. Within the women's beliefs, those in rural areas perceive pregnancy as a health status rather than a risky event. In respect of the economic aspects, the poor road infrastructure and the shortage of public transportation make it difficult to commute to health centers. Furthermore, Finlayson and Downe [113] reported that when pregnant women make an effort to attend prenatal care, the low-quality and unprofessional service provided by the health workers discourage the continuation of the perinatal care program.

Given the relevance of the service quality provided by the antenatal care providers, Simkhada *et al.*[334] recommend training midwives and nurses to be aware of the barriers women have to access to antenatal care, thus being sensitive to the women's

beliefs and socioeconomic situation. This awareness would be able to crucial to improve antenatal care services and expand its coverage in LMICs. Otherwise, antenatal coverage will be remained affected by factors such as the pregnant woman and husband education, pregnant woman's age, household income, availability of access to medical centers, and quality of perinatal care service [334].

The most causes of global perinatal death are preterm birth complications (35%), intrapartum related events (24%), and sepsis (15%) [367]. Studies conducted in LMICs have reported similar patterns, being prematurity, birth asphyxia, maternal hypertensive disorders, and septicemia, the most common causes [221, 12]. Fetus and newborns are also affected by infections, including syphilis, malaria, and animal and vector-borne diseases, [144, 135].

At intrapartum, the most leading cause of death is asphyxia [204, 387, 385, 133]. Asphyxia is related to obstruction on the placental blood flow, which can be due to maternal preeclampsia, placental abruption, or umbilical cord accident. The high death rate associated with asphyxia is mainly to poor delivery management. Asphyxia signs can be identified using fetal heart rate monitoring [112]. Detection of asphyxia can reduce irreversible organ damage, and identify cases requiring rapid deliveries [135]. However, this procedure is not often provided in LMICs.

In the majority of the reported perinatal deaths, low birth weight (LBW) is a common factor, with a prevalence between 70-80 % of the deaths [221, 12, 202]. LBW could be a consequence of either preterm births (< 37 weeks) or small-for-gestational-age (SGA). However, in LMICs, around 60% of low birth weight newborns are SGA [205]. The main reason for SGA in LMICs is IUGR, which has a prevalence of between 9-11% in LMICs [89, 205]. It can appear during pregnancy as a consequence of maternal vascular problems, malnutrition, or placental malfunction [407].

One approach contributing to reducing perinatal deaths is fetal heart rate monitoring. Fetal monitoring allows identifying non-reactive states associated with IUGR,

hypoxia, or intrapartum asphyxia [112]. By identifying antepartum or intrapartum risks, timely interventions can be performed.

In LMICs, fetal heart rate monitoring is usually provided by CTG or intermittent auscultation (IA), being the latter the most available. Although CTG is the standard of care in HICs, Housseine *et al.*[155] in a review of 37 studies carried on in LMICs found that CTG did not improve fetal outcomes in comparison to IA and performing CTG resulted in more cesarean rates without any benefit. In fact, a large historical reduction in perinatal deaths occurred with IA using a fetoscope [387].

Within IA methods, the Pinard stethoscope is the most available in resource-constrained regions due to its low cost[170]. At the labor ward, midwives used a stethoscope to listen to fetal heart rate for ten minutes every half hour [222]. However, auscultation with the stethoscope may be uncomfortable for both patients and practitioners, as well as being unsatisfactory to hear fetal beats due to the environmental noise [222].

Hand-held Doppler devices are also simple and inexpensive to use in LMICs. The hand-held Doppler allows providers to quickly assess fetal heart rate, to able to detect more fetal abnormalities than with Pinard stethoscopes [222]. In fact, Plotkin *et al.*[292] compared the performance of Doppler and Pinard devices for fetal monitoring in intrapartum by reviewing 19 studies conducted in India and African LMICs. The comparison showed that Doppler accurately detected more fetal abnormalities than the Pinard stethoscope. However, there was no statistically significant improvement in perinatal outcomes when an abnormality was detected. The authors suggested that the lack of improvement was due to the poor clinical management and protocol referral of abnormality events. The revision also found that both patients and medical providers preferred Doppler devices than Pinard, thereby justifying the integration of Doppler in fetal monitoring protocols for LMICs.

In addition to CTG and IA methods, few previous works showed the limited

use of Doppler imaging in LMICs. Seffah and Adanu[324] reported that detecting maternal and fetal anomalies using Doppler imaging is not very common in developing countries. Even if the device is available, usually there is a shortage of specialized physicians to provide the monitoring, and therefore short courses are provided to the medical staff. However, the training courses are not as adequate, standardized, and rigorous as those provided in HICs, thus challenging the use of ultrasound imaging [193]. Moreover, the Doppler imaging services are also affected by the frequent power outages, which affect the life span of the machines. Given the high cost of ultrasound imaging machines and the lack of evidence in improving perinatal outcomes at HICs, previous works have made questionable its use in developing countries [127, 126].

The use of fetal heart rate monitoring, coupled with cesareans, has shown the potential to prevent stillbirth and perinatal deaths [387]. In fact, according to Goldenberg *et al.*[134], any LMICs can reduce perinatal deaths by developing a prenatal care system to perform fetal heart rate monitoring, cesareans, if needed, or/and refer patients to advanced hospitals when complications are identified. However, the development of high-quality medical centers is a long-term solution requiring political will to make a significant investment in resources and facilities[411].

In the meanwhile, other alternatives, taking into account the LMICs' socio-economic situation, are needed. According to Drazancic[100], one of the most significant contributors to reduce perinatal deaths is the number of antenatal care visits to assess fetal well-being. Although increasing the number of antenatal visits is challenging in LMICs, particularly in rural areas, expanding mobile technology can help to overcome this problem [44].

2.9.1 Mobile technology for perinatal care

In recent years, mobile health applications have been developed to provide maternal and fetal monitoring. These applications are possible due to high penetration of

mobile telecommunications in LMICs, in which around 90% of population has a mobile device [44]. This high coverage of mobiles can be used to overcome perinatal care access barriers that LMICs suffers, such as low literacy level, poor road infrastructure, and lack of medical professionals and equipment [107, 342, 145].

The relevance of mHealth mobile apps to improve antenatal care was presented by Feroz *et al.*[110] 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 perinatal visits recommended by the World Health Organization (WHO). Feroz *et al.*[110] remarked that the most effective mobile apps used client education and behaviour change communication via SMS, as well as patient tracking that allows community health workers to follow-up on subsequent visits.

Relevant mobile phone solutions help community healthcare volunteers (CHV) to identify risks. Maitra and Kuntagod[223] developed an Android-based mobile application to help CHV to detect maternal mortality risks in an Indian community. The risk was calculated based on 35 questions ranging 12 different domains, such as demography, alcohol or drug usage, and past medical records. When a risk was identified, the patient was referred to a hospital for professional treatment. This app showed to be effective for assessing patients' health risk and reducing hospital workload. Similarly, an Android-based mobile app developed by Bakibinga *et al.*[31] guided CHV to identify maternal health risk in a sub county of Nairobi, Kenya. The mHealth application used WHO specifications to identify risks, thus referring a patient to a medical facility when needed. The patient's information sent from the mobile app to the hospital allowed them to easily identify referrals and provide timely treatment.

Mobile application can also be used for providing fetal monitoring. Although the initial implementation phases of these apps are usually developed in high-income

countries for home monitoring care, the majority of them have the potential to be used in LMICs. In a review of telemonitoring in obstetrics, Alves *et al.*[13] found that 92% of the reviewed mobile apps performed maternal and fetal monitoring by connecting the cellphone to external devices, such as electrodes, body sensors, and thermometers. External devices for maternal monitoring measured blood pressure, uterine contractions, changes in urine, and heart and respiratory rate. For fetal monitoring, on the other hand, external devices measured fetal heart rate. The used external devices allowed the digitalization of data, which later can be analyzed by medical professionals or artificial intelligence techniques to detect abnormalities. These mobile applications for fetal monitoring can focus on developing algorithms to detect fetal measurements from ultrasound images, as well as prototypes to provide fetal monitoring in an affordable manner. Within the algorithmic solutions, Jatmiko *et al.*[169] developed methods for smartphones to localize fetus from an ultrasound image. The algorithms were tested with images taken in a public hospital in Indonesia, achieving an accuracy of 93% for detecting fetal head and abdomen. Similarly, Khan *et al.*[183] developed algorithms to calculate the mean abdominal diameter (MAD) from an ultrasound image. The mobile app prototype was tested with ultrasound images recorded by professional midwives in a Norwegian hospital, resulting in a correlation coefficient of 0.96, mean error of -0.06 mm and a 95% confidence interval of -14.80 to 14.68 mm.

In respect of fetal monitoring prototypes, Awiti *et al.*[27] developed an android based digital fetoscope. The fetal heart sounds are acquired using a Pinard horn and a microphone, and are sent to a smartphone via Bluetooth. In the smathphone, the audio signals are processed to display the the heartbeat. The system is in a prototype phase and was tested on adults, comparing the measurements with those provided by a standard electronic sphygmomanometer. The android based digital fetoscope achieved a root mean square error of 7.23 beats with a standard deviation of 5.44 beats.

Tapia *et al.*[358] introduced a maternal-fetal monitoring in a resource-poor and educationally challenged community in Mexico. The project aimed to evaluate the feasibility to provide remote antenatal care. To that end, staff of a rural medical center were trained to use a mobile fetal monitoring composed of a fetal ultrasound heart monitor and a uterine tocodynamometer, as well as to use additional tools for recording maternal blood pressure, blood glucose, and urinary protein values. One-hundred fifty-five volunteers were equally split 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. Tapia *et al.*[358] observed that volunteers using the mHealth system were more than twice as likely than those using the standard of care to adhere to the antenatal care monitoring. Moreover, there were not any statistically significant difference in adverse perinatal outcomes between the two groups, thus suggesting that the prototype technology can be use in the rural community without compromise maternal and fetal health.

Stroux *et al.*[354] also introduced a fetal monitoring prototype in a Guatemala rural community. The introduced mHealth system consists of a low-cost Doppler transducer and blood pressure device connected to a smartphone running an app designed for low-literacy traditional birth attendants (TBA). TBAs were trained to use the mHealth system for home visits. When a TBA visited a patient, the app guided the TBA to find the fetus and record a Doppler recording of the heart beats. The app also guided her through basic questions presented through appropriate pictures and audio prompts to identify concerning signs and symptoms during pregnancy. In case a risk was identified, the app connected the TBA to appropriate medical care, either locally or remotely, to provide decision support and onward referral to appropriate healthcare if needed. The first release cycles showed that mHealth monitoring system is feasible to be used in the Guatemala rural areas [232, 231]

2.10 Discussion and Conclusion

Fetal monitoring is performed with different techniques, being CTG and ultrasound imaging the gold standard in high-income countries. Although, up to now, there is little evidence supporting the utility of these techniques to reduce perinatal mortality [92, 261], their use is still beneficial throughout pregnancy. Specifically, CTG can detect signs of hypoxia requiring cesarean delivery [140]. Ultrasound scan, on the other hand, used at early pregnancy (≤ 24) can help to date the gestation properly, and detect multiple pregnancies and fetal malformations [392, 141].

However, in resource-constrained settings, these techniques are scarce due to their high cost and requirement of specialized operators [100]. Moreover, socioeconomic, religious, and political aspects complicate the access of perinatal care to pregnant women in LMICs [113, 334]. All these factors have contributed to a high perinatal death rate in LMICs.

Experts recommend reducing perinatal death in LMICs using similar approaches to HICs, which had similar rates in the 30's [134]. Therefore, experts recommend organizing the health system in a regionalized network, in which patients can receive fetal monitoring, and provide cesarean when needed. However, this solution requires political will and costly investment, making it an unfeasible solution in the short-term [411].

One of the relevant indicators to reduce perinatal deaths is to increase the number of antenatal care visits [402]. An affordable solution to that aim is to use mobile technology and telemedicine [44]. Mobile technology allows bringing perinatal care services to rural communities, avoiding long commutes complicated by poor road infrastructure. The mobile applications can provide fetal monitoring by the cellphone to Doppler transducers, which have shown to be comparable to CTG for providing fetal heart rate monitoring in LMICs [155, 92], and which are widely available in those regions [37]. Moreover, as shown in Table 2.1, Doppler transducers have a no-

tably balance among cost, evidence, and time required for training operators, thereby making its use feasible. In addition to fetal monitoring, mobile applications can also record maternal symptoms and maternal recordings, such as blood pressure, glucose level, and urine protein levels.

The feasibility of these mHealth mobile applications has been shown in previous work conducted in LMICs [355, 110, 358]. The mHealth mobile applications emerged as a decision support system, are able to process recorded maternal and fetal information either by artificial intelligent functionalities or by sending the information to a centralized medical center via the cloud to be analyzed by trained medical staff. The feedback provided by the mHealth system may help to identify high-risk pregnant patients who need to be referred to principal public hospitals for further treatment. Therefore, the mHealth systems have the potential to provide perinatal care remotely, removing many barriers that pregnant women have to deal with in LMICs, and helping to avoid possible fetal complications.

Chapter 3

Signal quality method for assessing 1D-DUS in a clinical environment

3.1 Abstract

One dimensional Doppler Ultrasound (DUS) is a low cost method for fetal auscultation. However, accuracy of any metrics derived from the DUS signals depends on their quality, which relies heavily on operator skills. In low resource settings, where skill levels are sparse, it is important for the device to provide real time signal quality feedback to allow the re-recording of data. Retrospectively, signal quality assessment can help remove low quality recordings when processing large amounts of data. To this end, we proposed a novel template-based method, to assess DUS signal quality. Data used in this study were collected from 17 pregnant women using a low-cost transducer connected to a smart phone. Recordings were split into 1990 segments of 3.75 seconds duration, and hand labeled for quality by three independent annotators. The proposed template-based method uses Empirical Mode Decomposition (EMD) to allow detection of the fetal heart beats and segmentation into short, time-aligned temporal windows. Templates were derived for each 15-second window of the record-

ings. The DUS signal quality index (SQI) was calculated by correlating the segments in each window with the corresponding running template using four different preprocessing steps: (i) no additional preprocessing, (ii) linear resampling of each beat, (iii) dynamic time warping (DTW) of each beat and (iv) weighted DTW of each beat. The template-based SQIs were combined with additional features based on sample entropy and power spectral density. To assess the performance of the method, the dataset was split into training and test subsets. The training set was used to obtain the best combination of features for predicting the DUS quality using cross validation, and the test set was used to estimate the classification accuracy using bootstrap resampling. A median out of sample classification accuracy on the test set of 85.8% was found using three features; template-based SQI, sample entropy and the relative power in the 160 Hz to 660 Hz range. The results suggest that the new automated method can reliably assess the DUS quality, thereby helping users to consistently record DUS signals with acceptable quality for fetal monitoring.

3.2 Introduction

Although medical care has reduced mortality rates across the globe, birth has still remained an event of extreme risk. Approximately 2.6 million stillbirths and 2.8 million early neonatal deaths occur each year [401, 400]. Different factors contribute to this high burden, such as the lack of specialized medical professionals and the high cost of the medical devices, mainly affecting low and middle income countries (LMICs). Leading causes for perinatal mortality include Intrauterine Growth Restriction (IUGR) and congenital abnormalities of which, Congenital Heart Disease (CHD) is the most common [125, 203, 377]. These abnormalities are currently being detected using ultrasound imaging and more specifically, fetal echocardiography is performed for CHD diagnosis. However, these techniques are expensive and can only

be performed by trained sonographers or physicians; hence, their use is limited in LMICs [243].

Due to the high incidence and fatal consequences of these abnormalities in low-resource settings, affordable perinatal monitoring solutions are required. One of the most widely used, yet affordable methods for perinatal screening is fetal heart rate (FHR) monitoring. This technique has contributed to reduce perinatal and maternal risks through identification of non-reassuring fetal status [29]. Moreover, FHR has the potential for detecting IUGR [111, 267], as well as CHD complications [43, 82].

FHR monitoring is commonly performed through Cardiotocography (CTG) based on one dimensional Doppler Ultrasound (DUS), that is also used in low cost (under \$17) hand-held devices which can be operated by nonexperts. This DUS-based low-cost device has been used to develop affordable perinatal monitoring systems, thus facilitating screening in LMICs. Stroux *et al.* (2016) introduced a mobile-health monitoring system, based on a low-cost transducer and operated by illiterate birth attendants, to detect fetal compromise, such as IUGR, in rural Guatemala [354, 351]. DUS can also provide more information beyond the FHR, such as the cardiac valve function, which can further facilitate detection of CHDs and assessment of the fetal development [234, 235].

Despite the benefits of 1D-DUS, it is susceptible to noise affecting its quality, and it is non-stationary due to the fetal movements, which can complicate the FHR monitoring. Since the quality of the recorded signals is critical to properly detect FHR abnormalities, the assessment of the signal quality is an essential part of the recording process. Stroux and Clifford reported that the accuracy of FHR analysis depends on the signal quality, hence the quality should be ensured during the data collection [353]. Magenes *et al.* (2001) also showed the necessity of removing CTG signals with low quality before applying methods for detecting fetal anomalies [220]. The quality assessment process, enables providing feedback to the operator during

data collection, allowing them to retake or exclude the low-quality signals.

To date, little work has been published concerning the quality assessment of the DUS signals. Stroux and Clifford proposed a method to validate the quality of DUS signals recorded using a hand-held device connected to a smart phone [353, 351]. For this purpose, they extracted features based on sample entropy, wavelet decomposition coefficients, and the phone’s triaxial accelerometer output. To assess the quality, a logistic regression and a support vector machine (SVM) were trained to classify the recordings into noisy and clean categories. The logistic regression model was able to classify the signal quality with an accuracy of 95.14% on test data, while the SVM achieved an accuracy of 94.44%. Marzbanrad *et al.* (2015) proposed an automated method to assess the DUS signal quality for the application of fetal valve motion identification [238, 234]. In their method, DUS signals were segmented into cardiac cycles using non-invasive fetal electrocardiogram (fECG) as reference. Then, 12 features including power, statistical and entropy-based measures, were extracted from a frequency range associated with the fetal cardiac valve motion. Using these features, the signals collected from 57 fetuses were classified as good and poor quality, using a naive Bayes model. The accuracy of the classification was 86% using 10-fold cross validation.

In the current paper, to improve the quality assessment for perinatal monitoring, we propose a simpler template-based method using only the DUS signal recorded by a low cost device, thus facilitating its implementation in LMICs.

3.3 Methods

3.3.1 Database

The DUS database used in this paper was collected at the John Radcliffe Hospital in Oxford as a part of the study presented in [353, 351]. 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. This database contained 1D-DUS signals recorded from 17 subjects at a sampling frequency of 44.1 KHz using a hand held transducer (AngelSounds Fetal Heart Detector, Jumper Medical Co., Ltd.) with an ultrasound frequency of 3.3 MHz. Subjects were women with singleton pregnancy, over the age of 18, who were scheduled for a routine CTG. The duration of recording per subject is shown in Figure 3.1.

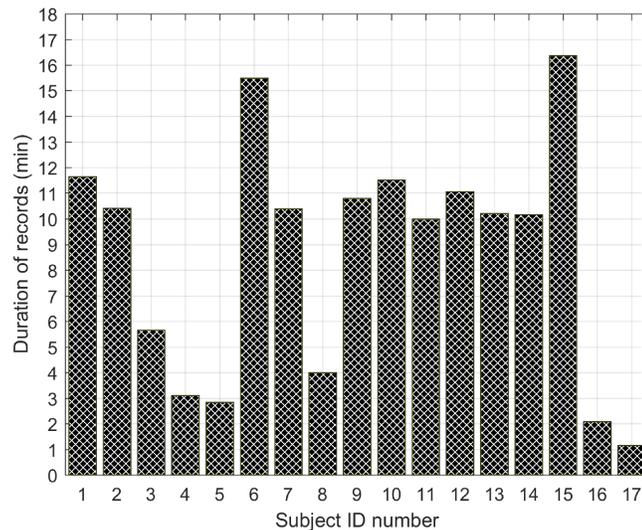


Figure 3.1: Duration in minutes of the total number of records per subject.

3.3.2 Segment selection

Each of the 1 minute-length DUS signals were labeled by three different annotators who had relevant experience in analyzing cardiac audio recordings for quality. Each reviewer independently labeled each second of the record as good or poor quality. After labeling, each record was split into segments of 3.75 with a 3 second sliding window (i.e. a 0.75 second overlap). The duration of 3.75 was fixed since it is the usual length for computerized analysis of fetal non-stress tests based on the Dawes/Redman

criteria [279, 85]. To ensure that each segment belongs to only one class, only the segments with all their samples of the same class were kept. These segments were assigned to 4 different classes as follows:

- Good Quality: Three annotators labeled all the segment as good quality.
- Mostly clean: Two annotators labeled all the segment as good and one labeled all the segment as poor quality.
- Mostly noisy: One annotator labeled all the segment as good and two labeled all the segment as poor quality.
- Poor Quality: Three annotators labeled all the segment poor quality.

A total of 1,990 segments (430 good, 1062 poor, 292 mostly clean, and 206 mostly noisy quality) were identified. Figure 3.2 illustrates the balance of segments across patients. Note that the quality of the recorded signals varies from one patient to another and may change over a single recording session because we observed that for some recordings there are both good and bad quality segments.

The classifier in this work was only trained using poor and good quality segments. The rationale behind this stems from the fact that segments on which one or more experts cannot agree are not meaningful in reporting statistics, since we cannot categorize them in a single class. However, after training the classifiers, the optimal classifier was also evaluated with the mostly clean and mostly noisy segments to determine its capacity for detect intermediate quality segments.

3.3.3 Preprocessing

The DUS signals were resampled at 4000 Hz using a least-square linear phase anti-aliasing filter. This downsampling does not affect the information content of the signal, since the fetal heart activity corresponds to the DUS signal frequencies below

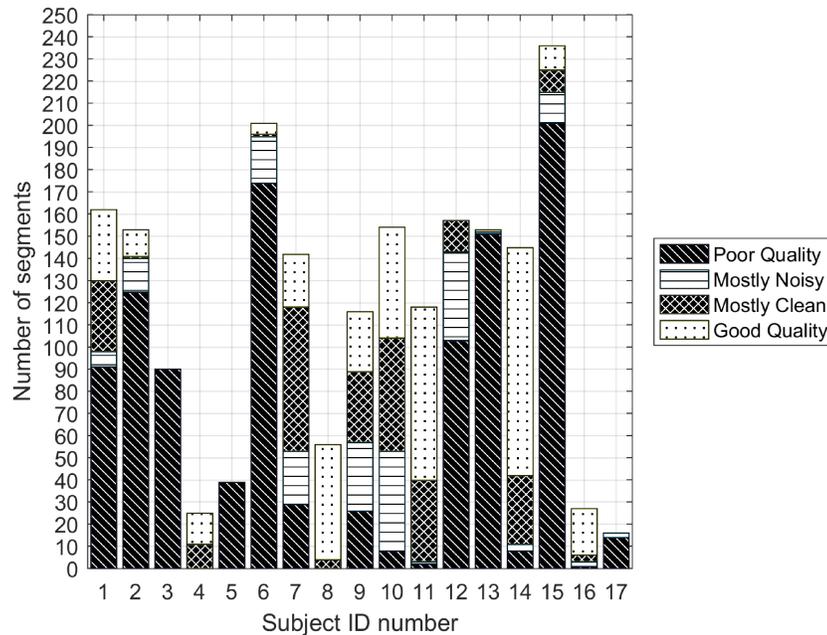


Figure 3.2: Number of poor and good quality 3.75 s segments for each of the subjects for which all three annotators agreed on labels.

1650 Hz for a transducer of 3.3 MHz [327]. Hence, the Nyquist-Shannon sampling criterion was satisfied after downsampling.

3.3.4 Template-based quality assessment of 1-D Doppler Ultrasound

To assess the quality of the DUS segment, a template-based algorithm was developed. This method consists of 4 stages (Figure 3.3). First, the beats of the DUS signals were estimated using Empirical Mode Decomposition (EMD). Then, using the estimated beats, templates for windows of 15 seconds were derived. These templates were then optimized in stage 3, and finally, the quality index of the DUS segment was calculated in stage 4. These stages are illustrated in Figure 3.3 and explained in the following sections.

This method, as all the remaining methods of this work, were implemented in Matlab and executed in a machine with a standard processor (Intel(R) Xeon(R)

CPU E5-2660 v2 @ 2.20GHz).

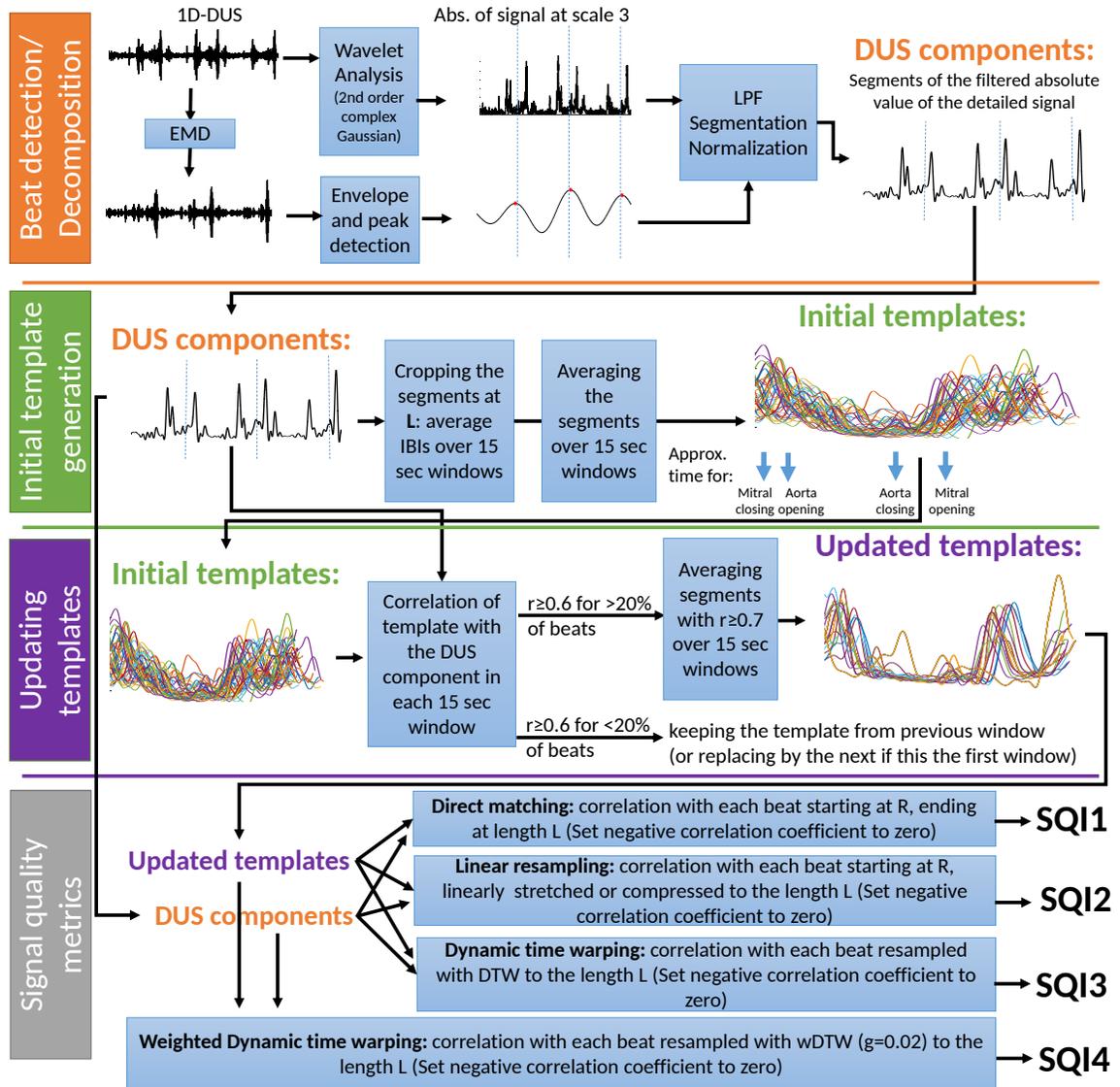


Figure 3.3: Overview of the template-based quality method for the Doppler signal

Beat Detection

Individual cardiac cycles (or beats) were detected using EMD, based on a method presented in [240]. EMD is an empirical method for decomposing non-stationary and non-linear signals into a set of components called Intrinsic Mode Functions (IMFs). It is a data-driven method that is able to adapt to the signal properties without

requiring a basis function, unlike other time-frequency decomposition methods [158]. This characteristic allows EMD to properly analyze non-linear and non-stationary natural processes.

Each extracted IMF satisfies 2 properties: firstly, the number of maxima and minima and the number of zero crossing should differ at most by 1; secondly, the mean value between the envelope of the local maxima and the envelope of the local minima must be zero at any point. To obtain the IMFs, EMD uses an intuitive algorithm called "sifting procedure". It is an iterative procedure, which finds all the IMFs of the signal until the difference between output and the input of the sifting procedure becomes a monotonic function. More details of the method can be found in [159].

To find the beats from the DUS signals an algorithm was developed to allow switching between the first four IMFs, which were obtained over 4 second windows. For each of these IMFs, the peaks were detected based on the positive first derivative and negative second derivative criteria. Then, using the identified peaks, the IMFs were enveloped to obtain four IMF envelopes for each window. To detect the best envelope for segmentation, a metric based on the standard deviation of the peak to peak intervals (PPIs) was applied. Namely, the IMF with the minimum average of the standard deviation of PPIs was selected as the optimum IMF. To deal with the possible mismatching of the selected IMFs in adjacent windows, a short overlapping window of 1 second was used to correct missing or double identified peaks. The peaks of the optimum IMFs were selected as possible beat locations. These peak locations were further corrected through a moving windows of 5 PPIs, replacing the middle PPI by the average of the rest in the window, if they differed by more than 20% [236]. The corrected peaks were set as beat location, and were used to segment the 1D-DUS signal into Beat to Beat Intervals (BTBI).

Continuous Wavelet transform (CWT) was then applied to the DUS signal over

25 second windows. In this work, the CWT was applied using second order complex Gaussian function as Wavelet mother. Moreover, only the signal decomposed at scale 3 was selected since it was found that the 3rd level is the most relevant for detecting value movement. This scale contained frequencies below 1000 Hz, which mainly reflect the fetal heart activity; valve motion is around 990Hz for a transducer of 3.3 MHz [256], and wall velocities between 257 and 429 Hz for a transducer of 3.3 MHz [327].

After applying the CWT, the envelope of the absolute value of the decomposed signal was estimated by interpolating the maxima. This envelope was then smoothed using a low pass filter and segmented into cardiac cycles using the estimated beat locations. Each segment was normalized by subtracting its mean and dividing by its standard deviation. These normalized segment were used to generate the templates for the signal quality assessment.

Initial template generation

Using the normalized cardiac cycle segments, the initial templates were calculated using a window of 15 seconds. The length (L) of the template was calculated as the average of BTBIs in each 15-second window. The initial template of the window was determined by averaging the segments starting at the beat location and lasting at length L. This procedure was repeated for each window, thereby obtaining an initial template for each window of 15 seconds.

Updating templates

The initial template of each window was updated based on the correlation function. For each window, the correlation of the template and the segments starting in a beat location and lasting at length L was calculated. The window template was updated by averaging only the segments with a correlation (r) greater than or equal to 0.6. In case the initial template of a window did not have a correlation of $r \geq 0.6$ for at least

20% of the beats, the template was assumed as invalid, and replaced by the one from the previous window. If the initial template of the previous window was invalid, the one from the next window was selected.

Signal quality metrics

After updating templates for each window, the quality indices were calculated as the correlation of the segments with the template in their corresponding window. The correlations was calculated in four ways:

- **Direct matching SQI.** The segments of each 15-second window, beginning at the beat location and ending at the length of the template (L) were used to calculate the correlation coefficient with the template and this was denoted as SQI1. If the segment was shorter than L , it was padded by zeros.
- **Linear resampling SQI.** Each estimated beat of the window was linearly stretched or compressed, if the length of the beat was shorter or longer than L , respectively. Then, the correlation coefficient with the template was denoted as SQI2.
- **Dynamic time warping SQI.** Using Dynamic Time Warping (DTW), the segments were transformed to the length L (as performed in our earlier work [209]) and the correlation with templates was denoted as SQI3.
- **Weighted dynamic time warping SQI** One drawback of DTW is that it gives too much freedom to the segment to adapt to the template. This was addressed by Jeong *et al.* (2011) who introduced the weighted DTW (wDTW) [172]. This method penalizes points with higher phase from the reference template by applying weights, thereby minimizing the distortion caused by outliers. In the current work, the parameter controlling the penalty was optimized through cross-validation to achieve the highest accuracy. The best value was

found to be 0.02. The correlation of the transformed segments in the window with the corresponding template, was denoted SQI4.

For all methods, any negative values of these SQI (negative correlation) were set to zero.

3.3.5 Sample Entropy and Power Spectrum Density (PSD)

In addition to the Template-based SQIs, two other key features were estimated from the DUS signals. The first one was sample entropy (H_s), which has shown a promising potential for discriminating between good and poor quality DUS segments [353]. Sample entropy measures the regularity of a signal by finding reoccurring patterns in it. To this end, three parameters are defined: the length of the signal N , the pattern length m and the matching tolerance r . Sample entropy is defined as the negative logarithm of the probability that a time series of length N with reoccurring pattern of length m within a set tolerance of r , also has reoccurring patterns of length $m + 1$ under the same tolerance constraint. In this work, the sample entropy was calculated setting the parameters $m = 2$, and r as 0.1 times the standard deviation of the input time series. The entropy was calculated using the procedure described in [303].

The second additional feature extracted was the Power Spectrum Density (PSD) ratio. This feature was used in order to evaluate the power of the DUS signals at different frequency ranges. The range for calculating the ratio was determined using a grid search. Since cardiac movements are associated with a Doppler frequency range of 100 Hz to 600 Hz using a 2MHz transducer [391], which translate to a scaled range of 165 Hz to 990 Hz for 3.3 MHz transducer, the ranges of values of the grid search were fixed from 80 Hz to 400 Hz and from 580 Hz to 900 Hz for the low and high frequency interval limits, respectively. For each possible pair of values, the capacity for discriminating between good and poor quality segments was measured using the earth mover's distance. The range with the highest earth mover's distance between

the distribution of the ratios of good and poor classes was found to fall in the range 160 Hz - 660 Hz. Thus, the PSD ratio of each DUS segment was calculated by dividing the power spectrum contained in the interval $[160 \text{ Hz} - 660 \text{ Hz}]$ by the total power, thereby measuring the percentage of power associated with cardiac movements.

3.3.6 Feature Vectors

Applying the template-based method resulted in four different SQIs for each estimated beat of the segment, thus obtaining a total of $4N_b$ indices by segment, where N_b is the number of beats of the segment. As the number of beats varied for each segment, we selected the median value of each quality index of the segment as the final SQI. Thus each segment had only one value for SQI1, SQI2, SQI3, and SQI4. Finally, the sample entropy and the PSD ratio were added to the feature vector, thereby obtaining a total of six features for each segment.

3.3.7 Classification

The above features were used to train an SVM classifier. SVM is a classifier that finds the best hyperplane that maximizes the margin between two classes. When the data are not linearly separable, a kernel function is used to transform the data to a different space in which the data can be separated. In this work, a Gaussian radial basis function kernel was used:

$$k(x_i, x_j) = \exp\left(\frac{-|x_i - x_j|^2}{2\sigma^2}\right), \quad (3.1)$$

where x_i and x_j are feature vectors, and σ is a free parameter of the kernel. A large value of σ increases the bias but reduces the variance of the classifier and a small value causes the opposite effect. To find the best value for a given training set, σ is usually tuned using heuristic methods or a brute force search.

The class prediction, y , of a given feature vector, x , is calculated using the dual representation of SVM as:

$$y = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i k(x_i, x) + b \right), \quad (3.2)$$

where x_i is the i -th feature vector of the training set and $y_i = [-1, 1]$ is its class; $\alpha \geq 0$ are Lagrange coefficients obtained by quadratic optimization; b is the intercept of the margin; and $k(x_i, x_j)$ is the kernel function (eq. 3.1). The α coefficient is only greater than 0 for those points that are in the margin. These points are called support vectors. In addition to the parameters of equation 3.2, SVM has a hyperparameter called the soft margin constraint (C). This parameter regularizes the margin allowing the cost function to ignore some points to establish an adequate margin for the training set. More details concerning the SVM can be found in [1]. In this work, the SVM parameters C and σ were optimized using five-fold cross-validation and a grid search on the training set. The grid search was defined by $C \in \{2^{-3}, 2^{-1}, \dots, 2^5\}$ and $\sigma \in \{2^{-5}, 2^{-2}, \dots, 2^2\}$.

3.3.8 Method performance assessment

To assess the performance of the method proposed here, the dataset was split into two equal subsets; the training and test sets. The the training set was used to determine the combination of features most relevant for assessing the quality of DUS segments. The test set provides an assessment of the accuracy on an independent dataset.

To split the dataset into two equal subsets (training and test sets), the subjects were ranked based on their number of good segments in descending order. Then, the data of each of the subjects were alternately assigned to the subsets. In other words, the first subject's samples were assigned to the training subset, the second ones to the test subset, the third one to the training subset, and so on. As the number of

patients was odd (17), the samples of the last subject were assigned to the subset with the lowest number of segments.

The best combination of features was found by calculating the accuracy of all possible feature combinations on the training set. Since the dataset presented an imbalance among classes (Figure 3.2), the accuracy was calculated using stratified five-fold cross validation with bootstrapping. Specifically, the accuracy of each feature combination was determined as follows: subjects of the training set were split into 5 folds. For each fold, 120 signals (60 per class) samples from the subjects of the fold were randomly selected using sampling with replacement (bootstrapping). The selection was performed in proportion to the subjects' sample quantity in each fold. The rationale behind this validation process is that the bootstrap applied to the cross validation folds adjusts the class imbalances, which is a critical factor for SVM classifiers [69]. Moreover, as the cross validation did not assign samples of the same subject to different folds, it provided an unbiased accuracy estimation. To obtain a more reliable accuracy, the described validation process was repeated 100 times, assigning subjects into different folds at each repetition.

For each iteration of the five-fold cross validation, the training set was normalized by subtracting the mean of the respective feature vector and dividing by its standard deviation. The test set was normalized using both mean and standard deviation derived from the training data. The cross validation accuracy of each iteration was averaged by selecting the median of the 5 accuracy values of the folds. Likewise, the accuracy of the 100 repetitions was selected as the median of the 100 accuracy values. This procedure was performed for each combination of features.

In addition to the overall accuracy of the classifiers, the sensitivity and specificity were also estimated. Sensitivity was defined as the proportion of good quality segments properly classified, whereas specificity denoted the portion of poor quality segments correctly classified.

To determine the capacity of the method for predicting intermediate quality segments, a SVM classifier was trained with the good and poor segments of the test set using the most common parameters C and σ for the 100 bootstrap repetitions. Once the best combination of features was determined (maximising accuracy, then specificity), the classifier was fixed and assessed on the test using the same bootstrap cross-validation validation procedure used for the training set. Finally, the probability of belonging to good class was also estimated for the mostly noisy, and mostly clean segments without retraining to assess the performance of the classifier on all data.

3.4 Results

3.4.1 Feature Selection

Table 3.1 presents the median accuracy, sensitivity and specificity of the best combination of input features for up to 6 possible features. As can be seen, the classifier was able to classify the quality of a DUS segments with up to 85.8% accuracy using either the combination SQI2 and sample Entropy (H_S), or the combination SQI2, PSD ratio and H_S . The accuracy of these two combinations of features resulted in a statistically significant improvement over the use of only one feature, H_S ($p < 0.05$, one-sided Wilcoxon rank-sum test). The results for all possible combination of features are presented in the appendix A (Table A.1).

It can also be seen from Table 3.1 that the sensitivity tended to decrease with an increase in the number of features, whereas specificity steadily increase until three features were used. Since the combination of both two and three features leads to equivalent accuracy, the combination SQI2, PSD ratio and H_S was chosen to be evaluated on the test set, since this maximizes specificity, and reduces the chances that a poor quality segment is labeled as good quality.

Table 3.1: Median classification performance of the 100 five-fold cross validation balanced with bootstrapping. IQR indicates inter-quartile range; ‡ indicates a significant improvement (Wilcoxon rank-sum test, $P < 0.05$) of a given feature combination compared to using a combination with one less feature.

Feature Combination	Median Accuracy \pm IQR (%)	Median Sensitivity (%)	Median Specificity (%)
H_s	84.2 ± 5.8	100.0	78.3
SQI2, H_s ‡	85.8 ± 5.0	93.3	80.0
SQI2, PSD, H_s	85.8 ± 5.0	83.3	90.0
SQI2, SQI4, PSD, H_s	85.0 ± 8.3	85.8	88.3
SQI1, SQI2, SQI4, PSD, H_s	84.7 ± 5.0	85.0	86.7
SQI1, SQI2, SQI3, SQI4, PSD, H_s	83.8 ± 6.7	81.7	86.7

3.4.2 Test Set Performance

Table 3.2 displays the accuracy, sensitivity and specificity of the combination of the SQI2, PSD ratio and H_s features. The median accuracy of this classifier using this combination was similar to the highest median accuracy achieved on the training set. However, the interquartile range for the test set was almost twice than that for the training set, indicating that the test set may exhibit a higher heterogeneity of features. Both sensitivity and specificity exceeded 90%.

Table 3.2: Performance of the classifier averaged over 100 five-fold cross validation runs balanced with bootstrapping for the test set (with) using SQI2, PSD ratio, and sample entropy (SQI2, PSD, H_s) as features.

Measure	Minimum (%)	1st Quantile (%)	Median (%)	3rd Quantile (%)	Maximum (%)
Accuracy	65.8	79.2	85.8	90.0	96.7
Sensitivity	71.7	85.0	91.7	96.7	100.0
Specificity	61.7	89.3	91.7	95.0	98.3

3.4.3 Performance of classifier on intermediate quality segments

The classes of the mostly clean and mostly noise segments of the test set were also predicted using the same classifier (using SQI2, PSD ratio and sample entropy (H_s))

as features). Note that these segments were not used in training. Figure 3.4 shows the relative distribution of output probabilities from the classifier of belonging to the good quality class for all four types of segments. The classifier established a probability threshold of 0.5575 for distinguishing between good quality and poor quality segments. The percentage of segments which lay above the threshold for good quality and mostly clean segments were 86.53% and 69.06%, respectively. On the other hand, the percentage of segments that lay below the threshold for poor quality and mostly noise segments were 96.50% and 63.69%, respectively.

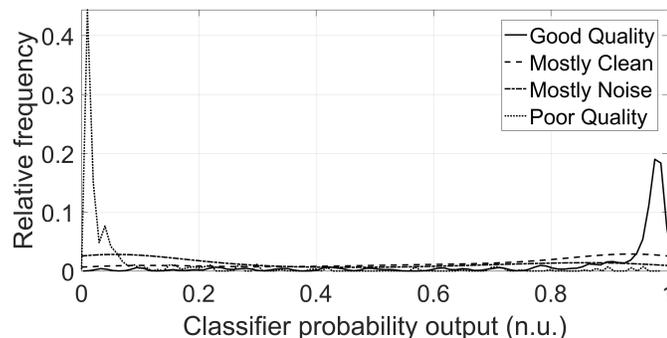


Figure 3.4: Distributions of classifier probability outputs for DUS segments of test set for each of the four classes (n.u. stands for normalized units). The threshold of belonging to the Good class was fixed at 0.56 for the classifier. The majority of the distribution of the Good and Mostly Clean classes lies above this threshold, whereas the majority of Poor and Mostly Noise classes lies below this threshold, as was expected. The probability distributions were smoothed using a normal kernel function [53].

3.5 Discussion

The results presented here suggest that it is possible to accurately classify the DUS quality using SQIs derived from DUS signals alone. Among the extracted features, sample entropy and PSD ratio provided suitable discrimination between good and poor quality segments, which is consistent with previous works [238, 353]. However, the addition of our proposed template-based method, particularly after linear resampling of the beats to match the running template (SQI2), provided a statistically significant improvement in accuracy (see Table 3.1). Either combinations SQI2 and

H_s or SQI2, PSD ratio and H_s resulted in a statistically significant accuracy; nevertheless, in order to maximize specificity, the combination of SQI2, PSD ratio and H_s was selected for assessing the classifier on the test set.

The selected features achieved an accuracy of 85.8% on the test set, thus suggesting that this metric is suitable for quality assessment based only on DUS signals. Although this feature combination exhibited more variance on the test set than on the train set, the achieved accuracy indicates that the model was not overfitted, and its complexity of three features is viable for assessing DUS quality. Furthermore, the balance towards specificity provided by the three chosen features (SQI2, PSD ratio, and sample entropy) ensures a high number of good quality segments is preserved, as well as small number of false-positive segments.

The best combination of features also showed an adequate capacity for classifying segments associated with intermediate quality zones (mostly noisy and moistly clean segments). Although both mostly clean segments and mostly noisy segments exhibited a mostly flat distribution, their centers were more closer to the good quality center and poor quality centers, respectively as it was expected. Specifically, almost 70% of the mostly clean probabilities laid above the SVM prediction threshold, whereas higher than 63% of the mostly noise probabilities laid below the SVM prediction threshold. This discrimination ability for the two ambiguous classes indicates the potential of the approach outlined in this work.

Regarding the template-based SQIs, the EMD-based approach appeared to facilitate identification of beat intervals, since the correlation with each template was generally high. Small offsets in the relative start and end point of each beat were mitigated by the use of resampling prior to correlation. The segmentation facilitated beat-by-beat quality assessment, which is the first step towards detecting fetal abnormalities from DUS signals. The template optimization process obtained representative templates for quality assessment since the initial template was only averaged

with those segments which exhibited a moderate or strong correlation ($r \geq 0.6$) with the initial template (average of first N beats).

Although Stroux and Clifford reported a higher accuracy (95.14%) on the same database [353], their work cannot directly compared to the current work since different statistical validation approaches were used. Specifically, Stroux and Clifford trained on two thirds of the data set and held out one third for testing, with no cross validation. In this work, stratified five-fold cross validation was used with bootstrapping (repeated 100 times), with subject stratification across different folds in each repetition. The accuracy obtained in this work cannot directly compared to that of Marzbanrad *et al.* (2015)[238] since they tested their method with a different dataset. However, our method can be compared to the aforementioned previous works by analyzing the effect of adding the index SQI2 to common features of the other works, namely, sample entropy. As was previously showed in Table 3.1, by using the SQI2 feature in addition to sample entropy, the accuracy statically significantly increased.

Another advantage of our method over previous works is that the proposed method does not need additional sources, such as accelerometer data [353] or an fECG signal [238] to assess the DUS quality. A key advantage of using only DUS signal is that the recording process is simple, facilitating the use of this technology by non-experts in low-resource settings. Finally, using only one source for quality assessment reduces health screening costs, facilitating its use in LMICs.

Despite the promising results, one limitation of the current method is that it was only tested using DUS signals recorded by professional midwives in hospital settings. The LMIC context is often severely resource constrained and there is a lack of widespread training for midwives, particularly in the use of technology. Consequently, signal quality is likely to be lower in recordings taken in LMICs. The noise content may also be different if the audio cable is not incorrectly inserted, introducing ambient sounds such as animals, extreme weather, and interference from non

hospital electronics. Nevertheless, the template-based method proposed here could be adjusted to specific conditions with a relevant training set.

Another possible limitation may be that the introduced method was only tested using one database labeled by three annotators. As DUS quality annotation is prone to inter-observer variability, testing the method with datasets annotated by different experts may reduce the accuracy. However, the high accuracy achieved by the combination of sample entropy, PSD ratio and SQI2 used in this work, provides optimism for the use of the template-based method for different datasets, especially with retraining.

3.6 Conclusion

The work presented in this article proposed a template-based method to assess the quality of 1D-DUS signals recorded by a low cost device. The introduced template-based indices provided a simpler method based on only DUS signals, thus facilitating its implementation in LMICs. The approach described in this work can provide the operator with an accurate and timely feedback on the quality of the recordings, to allow discarding the low quality signals in real time and prompt users to re-record data. Therefore, this quality assessment technique could potentially facilitate reliable fetal monitoring by non-experts towards reducing perinatal health burdens.

Chapter 4

Data capture errors in the 1D-DUS recorded in the field

4.1 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 in a manner that enables the user to take immediate corrective action.

In this work, we present an example of such a system, which provides real time feedback on the source of noise and interference on a low cost Doppler device connected to a smart-phone used by traditional birth attendants (TBAs) in rural Guatemala. A total of 195 fetal recordings made from 146 singleton pregnancies in the second and third trimester were recorded over 8 months by 19 TBAs. The resulting 33.7 hours of data were segmented into 0.75 s epochs and independently labeled by three trained researchers into one of five noise or quality categories that dominated the data.

A two-step classifier, composed of a logistic regression and a multiclass support

vector machine, was then trained to classify the data on epochs from 0.75 s to 3.75 s. After feature selection the highest micro-averaged test F1 score was 96.8% and macro-average F1 test score was 94.5% for 3.75 s segments using 23 features. A reduced real time model with 17 features produced comparable micro- and macro-averaged test F1 scores of 96.0% and 94.5% respectively.

The code is portable back to a low-end smartphone to run on such a device in real time (under 400 ms) in order to provide an audiovisual cue for the TBAs via the smartphone. Future work will evaluate the classifier presented here as part of a decision support system for data quality improvement in an ongoing randomized control trial in Guatemala.

4.2 Introduction

Low and middle income countries (LMICs) account for approximately 98% of all reported perinatal deaths world-wide [412]. Every year, around 2.6 million stillbirths and 2.8 million early neonatal deaths occur in these countries [401, 400]. One of the leading causes of these high mortality rates is Intrauterine Growth Restriction (IUGR) with a prevalence of around 11% in LMICs [48, 89]. IUGR is usually prevented in high income countries by performing routine perinatal screening and appropriate medical referral; however, this procedure is not appropriately followed in LMICs, due to the limited number of trained physicians and healthcare professionals and the high cost of medical equipment, particularly in rural areas.

One of the LMICs dealing with this challenge is Guatemala, where the perinatal morbidity and mortality rates are the highest in Latin America [404, 367]. This is particularly marked in the Maya indigenous communities where women face a range of barriers to access health services. Their pregnancies, if monitored at all, are likely to be attended by lay indigenous midwives, or traditional birth attendants (TBAs),

who often lack access to adequate equipment and training for identification of fetal health risks.

Recent 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 [351, 354, 232]. The system allows the monitoring of fetal heart rhythm, which is commonly used by clinicians to identify non-reassuring fetal status for timely intervention [29]. 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 TBAs were trained to use the approach for monitoring fetal well-being during pregnancy [351, 354, 232].

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 [352, 351, 354, 232], 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 [232]. Despite the quality improvements by retraining the birth attendants and fixing the device connections, the signal quality still needs 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 [353, 220].

Related work on signal quality can be found in the domain of heart sound analysis

[347] and other physiological monitoring such as the electrocardiogram [75, 39, 74], blood pressure waveform [210, 357] and photoplethysmogram [209]. Similar to these modalities, DUS signals are characterized by a high and fluctuating SNR ratio, as well as to be non-stationary due to the movement and angle of the probe against the skin. However, despite these issues, good quality DUS segments clearly exhibit a relative periodicity of fetal heart beats. Based on this periodicity, previous works have proposed to assess DUS quality by evaluating the presence of reoccurring signal patterns. Specifically, Stroux and Clifford *et al.* (2014) [353] first introduced a Signal Quality Index (SQI) using sample entropy, energy percentage of wavelet coefficients and the smartphone’s triaxial accelerometer to asses DUS recorded with same affordable 1D-DUS device used here. A logistic regression classifier was then trained to differentiate high from low quality expert-annotated data with an accuracy of 95.14%. Subsequently, Valderrama *et al.* (2017) [373] introduced a template-based method for segmenting DUS into their beat to beat intervals, calculating four SQIs for each beat. The SQIs were added to sample entropy and a power spectrum density ratio in the interval [160 Hz – 660 Hz], achieving an accuracy of 85% for classifying expert-annotated data. In addition to these works, a similar work for assessing quality of DUS recorded with a different transducer was performed by Marzbanrad *et al.* (2015) [238]. That work proposed an automated method to assess the DUS signal quality for a more specific application of fetal valve motion identification. To that end, 12 features including power, statistical and entropy-based features were extracted. The signals collected from 57 fetuses were classified as good and poor quality, using a naïve Bayes model, with 86% accuracy using 10-fold cross validation.

However, these earlier works were conducted on data recorded in hospital environments by skilled users and were not optimized to run on a smartphone. In the LMIC setting, noise and interference is likely to manifest in different ways, and generally present a more challenging problem. In preliminary analyses of the data used

in this study, substantial differences in noise content and errors in device use were discovered between the original hospital data and the field data. In particular, it was noticed that the following key issues were prevalent: 1) incorrect cable connections introducing ambient sounds such as those of animals, extreme weather and human speech, or just leading to a low volume/silent recording and 2) interference from electronics, such as mobile phones. Many of these problems cannot be mitigated with post processing and therefore a point-of-care system for suggesting ways to improve quality at the moment of data collection is imperative.

In this article, a system is presented for identifying the type of noise present in the DUS at the point of collection which is portable back to a smartphone to run in real time and provide an audiovisual cue for TBAs (Figure 4.1). This real time feedback would endow the mHealth monitoring system with a key crucial building block to discard low quality recordings and suggest re-recording of data, which is vital before identification of any fetal abnormality through 1D-DUS.

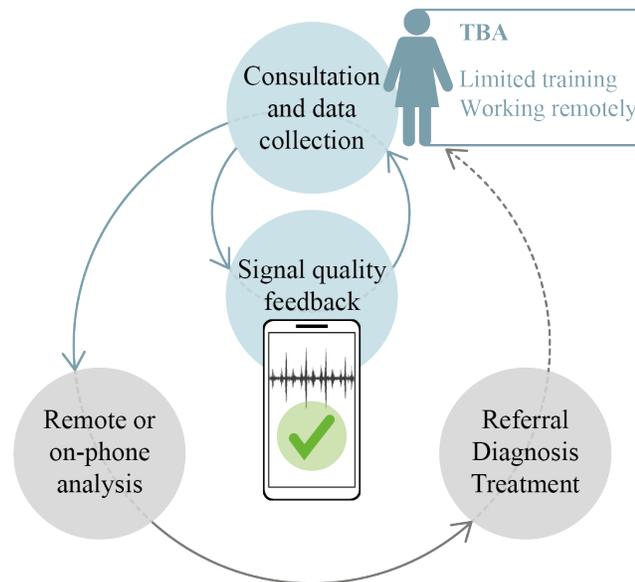


Figure 4.1: mHealth monitoring scenario. Providing traditional birth attendants (TBAs) with signal quality feedback at point of data collection is crucial for adequate referral, diagnosis and treatment.

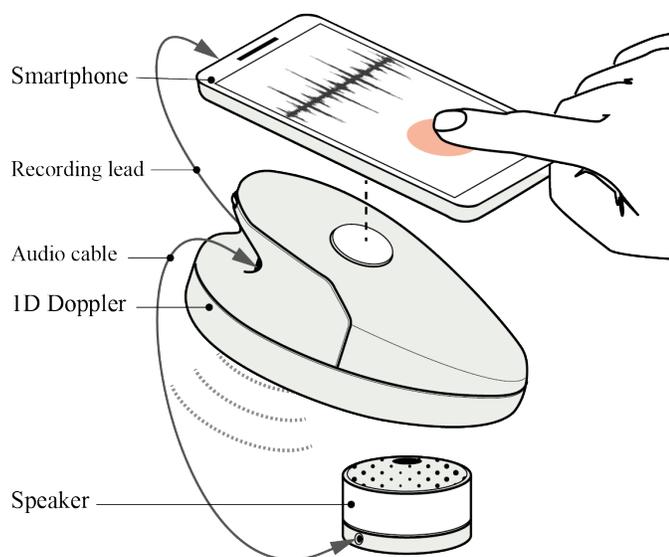


Figure 4.2: Illustration of the Doppler-mobile phone recorder used for data acquisition. The smartphone is completely audio- and pictogram-driven for low literacy populations. It attaches to an ultrasound device for recording, which in turn attaches to a low cost speaker so that the user has immediate audio feedback of the fetal heartbeat.

4.3 Methods

4.3.1 Data acquisition

The data collected in this work was part of a randomized controlled trial, conducted in rural highland Guatemala in the vicinity of Tecpán, Chimaltenango, between the 13th of May 2016 and the 20th of December 2016. 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). TBAs were trained to use a low cost (US\$17) 1D-DUS device, connected to a smartphone and external speaker (Figure 4.2). Additionally, the TBAs were given a self inflating blood pressure device (Omron M7) and bespoke software on the smartphone to guide them through the performance of a series of tasks. The app served several functions, including compiling responses to a checklist of serious issues, which would trigger contact with medical personnel at

appropriate points, capturing medical data and ultrasound recordings, and uploading data via SMS, GPRS and WiFi to Amazon Web Services. Each TBA was trained to use the Doppler device for a few hours over the course of a week, and had subsequent access to a project nurse for later support.

An agile approach was taken to the app design and implementation of the software, with programmers making periodic visits to the study sight in Guatemala. More details on the design and implementation of the data collection system, and the training of the TBAs can be found in Stroux *et al.* (2016) [354] and Martinez *et al.* (2017) [232].

4.3.2 Description of data

The DUS recordings used in this work were recorded by 19 different TBAs using a hand held DUS 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. A total of 146 mothers with singleton pregnancies were monitored, all indigenous Maya women between the second and third trimester of pregnancy. From these subjects, 195 DUS recordings were extracted. The median number of recordings per TBA was 9 with an interquartile range (IQR) of 9.25. More recordings were made in the third trimester (a median of 7 (IQR 5.75) versus 1 (IQR 3.75) in the second trimester. See Figure 4.3. The median duration of the recordings over all TBAs was 10.18 minutes with an IQR of 1.21 minutes.

4.3.3 Preprocessing

Each ultrasound recording was down-sampled to 4 kHz using a least-square linear phase anti-aliasing filter. This satisfied the Nyquist-Shannon sampling criteria as the information content of fetal heart activity in Doppler signals has been observed in frequency ranges up to 1 kHz with a 2 MHz transducer, which is equivalent to 1650

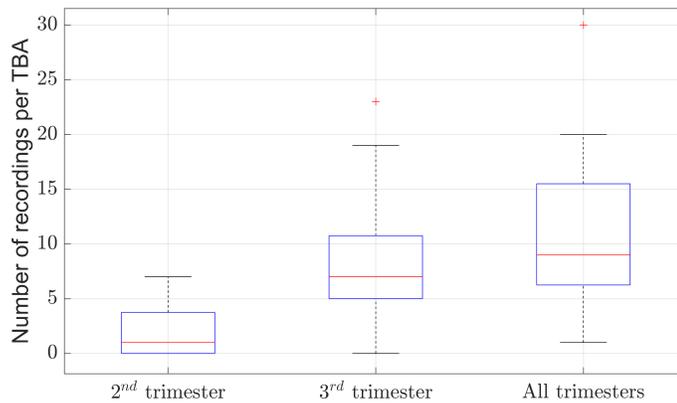


Figure 4.3: Box plots (median \pm IQR) of number of recordings by TBAs in the second and third trimesters, and over both the latter trimesters. By design, no recordings were performed in the first trimesters.

Hz for a 3.3 MHz device [327].

4.3.4 Class annotation

Three independent annotators listened, visually inspected and labeled all the 195 DUS recordings using a Matlab (MathWorks, Natick, MA, USA) graphical user-interface (GUI) interface (Figure 4.4). 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 short 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 to a specific class, or the annotator was unsure of which class it belonged to.

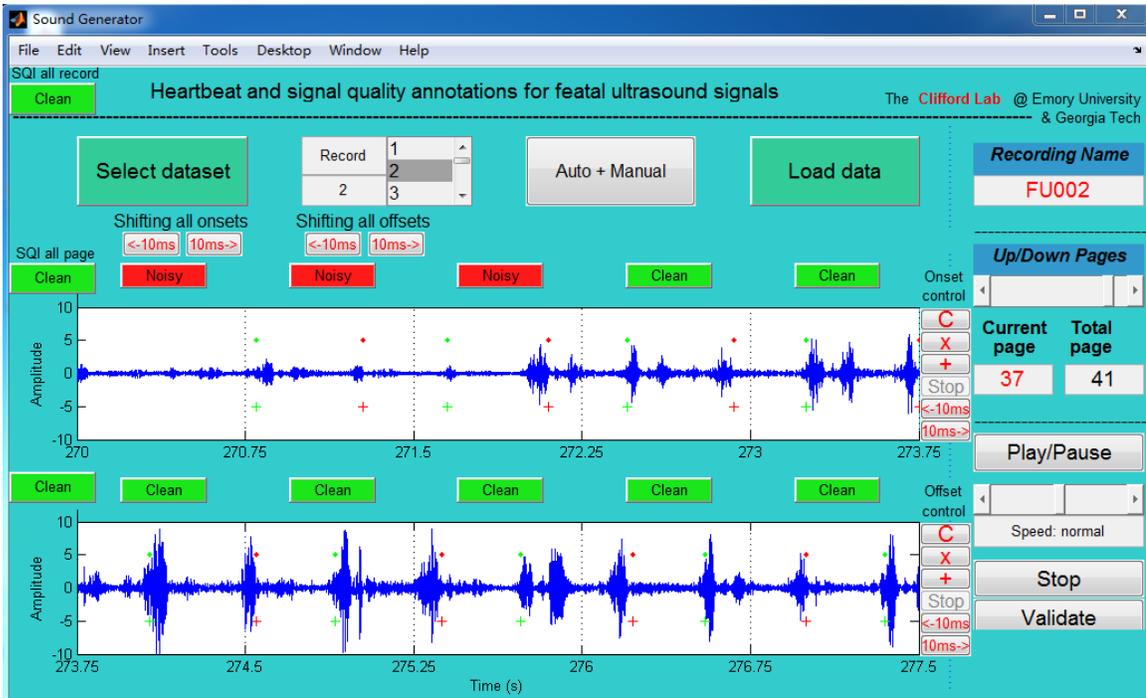


Figure 4.4: GUI used for assessing DUS recordings quality. The blue tracing represents two contiguous strips of 3.75 s raw audio file, each broken into five 0.75 s segments. The entire 7.5 s segment could be played back in real time or at fractional speeds (with pitch-preserving frequency shifting), paused or looped. The green and red crosses indicate the start and end of each ‘heart beat’ as determined by an automated algorithm [351], 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*.

- *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 [279, 85]. 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

Table 4.1: Number of segments per class for each window length after annotation, and adjudication.

Segment length (s)	<i>Good</i>	<i>Poor</i>	<i>Interference</i>	<i>Talking</i>	<i>Silent</i>	<i>Unsure</i>
0.75	26,090	14,738	2,832	8,718	18,580	2,356
1.50	25,074	13,062	2,478	7,744	18,261	1,871
2.25	24,086	11,775	2,163	6,966	17,993	1,589
3.00	23,147	10,813	1,888	6,409	17,775	1,376
3.75	22,263	10,008	1,647	5,974	17,579	1,200

appear more than once in a longer window. The number of segments obtained after keeping homogeneous segments are displayed in Table 4.1.

4.3.5 Feature extraction

To extend earlier work for binary classification (good vs. poor) of the 1D-DUS data, several new features were added, which were specific to the four subsets of noise observed in the data.

Template-based features

Following Valderrama *et al.* (2017) [373], four SQIs were calculated for each estimated beat to beat interval (BTBI). To estimate the BTBIs, the DUS signal was decomposed into its Instantaneous Mode Functions (IMFs) by applying Empirical Mode Decomposition (EMD) [158]. Then, the IMFs' envelopes were detected using their peaks, identified based on the positive first derivative and negative second derivative criteria. The IMF with the smallest average of standard deviation of the peak to peak intervals was used to find the beat locations from its peaks. These beat locations were used to segment the high frequency components of DUS corresponding to the fetal heart activity, which were found by using Continuous Wavelet Transform (CWT) with a mother Wavelet second order complex Gaussian function at scale three. This scale contained frequencies below 1000 Hz (pseudo-frequency), which mainly reflect the fetal heart activity; valve motion is around 990 Hz [255], and wall velocities be-

tween 257 and 429 Hz [327]. The higher frequency components of the the DUS were then segmented and normalized.

Initial templates for each window were calculated by averaging over all segmented beats in the first 15 s. The initial template was updated by averaging it with segmented beats that exhibited a correlation, $r > 0.6$, with the running template. Although this weights the current beat as much as all previous beats, the requirement to have a modest correlation with the running template, balances adaptation and preservation of information. This value of r was optimized in an earlier work on DUS recorded in a hospital environment with the same device [373]. Finally, using the optimal templates, four different SQI were calculated as:

- **SQI1: Direct matching SQI.** An optimal template was correlated with each DUS components and this was denoted as SQI1. If the DUS component was shorter than length of the optimal template (L), it was padded by zeros. On the other hand, if DUS component was longer, only the first L samples were correlated with the template.
- **SQI2: Linear resampling SQI.** Each DUS component was linearly stretched or compressed to the length of the optimal template. Then, the correlation coefficient with the template was denoted as SQI2 calculated.
- **SQI3: Dynamic time warping SQI.** Using Dynamic Time Warping (DTW), each DUS component was transformed to the length of the optimal template (as performed in Li *et al.* (2012)[209]) and then, the correlation with the optimal template was calculated.
- **SQI4: Weighted dynamic time warping SQI.** Due to the fact that DTW gives too much freedom to a segment to adapt to the target segment, DUS component was transformed using the weighted DTW (wDTW) [172]. This method penalized points with higher phase difference from the reference template by ap-

plying weights, thereby minimizing the distortion caused by outliers. To obtain an adequate penalty, cross-validation was performing into DUS recorded with the same device presented here, resulting in a best penalty value of 0.02 [373]. The correlation of the transformed DUS component with the optimal template was then calculated.

For all methods, any negative values of these SQI (negative correlation) are set to zeros.

Since this method extracts four SQI for each DUS component contained in a defined window length, the number of SQIs for a window length varies. The median value of each SQI was therefore calculated for each window. Thus, each window had only four SQIs (SQI1, SQI2, SQI3, and SQI4).

Sample entropy features

Sample entropy measures the regularity of a signal by finding reoccurring patterns. Three parameters need to be defined: signal length N , pattern length m and matching tolerance γ . Using these parameters, the sample entropy H_s is defined as the negative logarithm of the probability that a length N time series, that has repeated itself for data points of length m within set tolerance γ , will also repeat itself for $m + 1$ data points within the same tolerance constraints [303]. For this work, the parameter were set $m = 2$, and γ as 0.1 times the standard deviation of the input time series since these value obtained promising results for DUS signals recorded with the same device [353]. The entropy was calculated using the procedure described in Richman and Moorman (2000) [303].

Wavelet features

Wavelet analysis is often described as an alternative to the short-term Fourier transform and particularly suited to the analysis of localized power within non-stationary

Table 4.2: Wavelet frequency decomposition. Level 1 and 2 contain valve activities, reported to be in the range 990-3300 Hz for the ultrasound transducer used here [255]. Level 3 corresponds to wall velocities, which are located at range of 257-429 Hz, given the Doppler transducer used in this work [327]. Level 4 included frequency that is partly associated with fetal body movement [405]

Level	Frequency range (Hz)	Coefficients	Fetal event
1	1000-2000	d1	valve activities
2	500-1000	d2	valve activities
3	250-500	d3	wall velocities
4	125-250	d4	fetal body movement

signals, as it has simultaneously high temporal and frequency resolution [306]. The frequency resolution can be performed in the discrete domain by using multi-resolution analysis, in which the signal is filtered with high and low-pass filters to separate the frequency content into frequency bands of the same width. Each decomposition step halves the frequency bandwidth (doubles the frequency resolution) while its output is downsampled by a factor of two (thus halving the temporal resolution).

To select a sensible frequency range for DUS signals, the velocity ranges of the fetal myocardium were reviewed. Peak axial cardiac wall velocities have been measured between 60 and 100 mm/s [327], which translates to a frequency range of 257-429 Hz, given the Doppler transducer used in this work. Similarly, valve motions have been reported to be around higher frequencies [239, 296]. Specifically, Murata and Martin (1974) [255] described the use of a 600-2000 Hz bandpass filter to monitor valve activity, which translates to a 990-3300 Hz range for the ultrasound transducer used in this study. Using these frequency ranges as reference, the DUS recordings were decomposed into 4 levels using the Discrete Wavelet multi-resolution analysis (Table 4.2). The reverse biorthogonal wavelet *rbio3.9* was selected as the mother wavelet since it was able to correctly classify more good and poor quality segments than other mother wavelets for DUS recordings made with the same transducer used in this study [353, 351].

As a feature, the percentage of energy content at each decomposition level was

computed as follows:

$$E_d(k) = 100 \frac{\sum_{i=1}^{N_k} c_k(i)^2}{E_t}, \quad (4.1)$$

where k denotes the level of decomposition, $c_k(i)$ the i -th detail coefficient at level k , N_k the number coefficient at level k , and E_t the sum of the squared coefficients over all levels.

Statistical features

Two SQI's based on the DUS statistical characteristics were also calculated. The first SQI was the variance, measuring total power or by how the signal varied about its mean. This feature was included specifically to deal with the silent class, whose amplitude is relatively low, thereby creating vanishingly small variances.

The second SQI was calculated based on the autocorrelation function, which has been used for assessing quality of other types of cardiac sound signals [346, 192, 212, 211]. Specifically, to classify noise segments, the kurtosis of the autocorrelation was calculated. Noise segments tend to follow a Gaussian distribution, and therefore, it is expected that their kurtosis be close to 3. On the other hand, good quality signals are likely to have larger kurtosis values since they consist of large excursions with little time around the baseline, provoked by strong audible heart sounds.

Power spectrum density (features)

Following work presented in Valderrama *et al.* (2017) [373], a series of power spectral density (PSD) ratios associated with cardiac movements and key spectral contaminants were implemented. As noted, cardiac movements generally manifest within a Doppler frequency range of 100 Hz to 600 Hz when a 2 MHz transducer is used [391], which translates to a scaled range of 165 Hz to 990 Hz for 3.3 MHz transducer. Per earlier work [373], a ratio of the power contained in the interval [160 Hz – 660 Hz] over the power of the entire signal was included as key feature for identifying normal

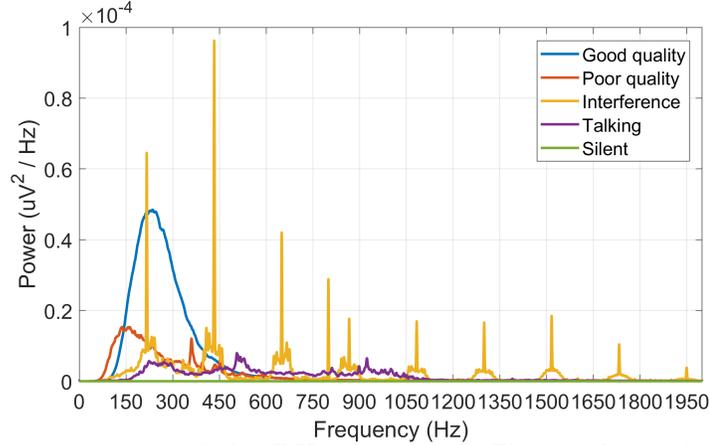


Figure 4.5: Power spectrum of the DUS segments. Electrical interference presents pronounced peaks around 217 and in its harmonics (i.e. 434 Hz, 651 Hz, 868 Hz, etc) which are associated to the radio protocol GSM-TDMA [21]. There is also a pronounced peak around 800 Hz, which is a frequency associated with audio distortion caused by amplifiers [4].

activity.

In addition to this cardiac movement PSD ratio, other PSD ratios were also included to detect interference segments. To define the frequency PSD intervals for electronic interference class, the average PSD of the DUS segments was calculated per class (Figure 4.5). Two different obvious spectral of peaks were identified in the interference class. The first key peak is associated with the GSM-TDMA radial protocol, which appears around 217 Hz and in its harmonics (i.e. 434, 651, 868 Hz, etc) [21]. The other pronounced energy peak was identified to be around 800 Hz, a frequency associated with audio distortion caused by amplifiers [4]. Based on these locations, three PSD ratios were calculated; two for the largest peaks of the GSM protocol, and the remaining for the amplifier distortion peak. These PSD ratios were calculated as:

$$P_r = \frac{\sum_{i=p_r-20}^{p_r+20} P_i}{\sum_{i=p_r-100}^{p_r+100} P_i}, \quad (4.2)$$

where p_r is one of the locations of the peaks (217 Hz, 434 Hz, and 800 Hz).

Speech detection features

Two types of features associated with speech were included: the estimation of fundamental frequency and Mel-Frequency Cepstral Coefficients (MFCCs). The value of F_0 was estimated using the cepstrum method. The real cepstrum was applied to each segment using framing of 25 ms with a hamming window and a sliding window of 10 ms (15 ms overlap), which are the most common window's length for speech recognition [160]. For each frame, the F_0 was estimated by taking the inverse of the quefrequency within 50 and 1000 Hz with the highest peak. After the F_0 was estimated for each possible window of the segment, all the values were distributed in a histogram, calculating the number of bins from the average provided by Sturgis' [356] and Rice's [197] rules. The F_0 of a segment was defined as the center of the histogram bin with the highest absolute frequency.

To calculate the MFCCs, a standard algorithm [408] was used, which is summarized as:

1. Accentuate the key spectral band of the signal using a first-order high pass filter with an α of 0.95.
2. Frame the window using a hamming window and a window length of 25 ms with a sliding window of 10 ms.
3. Calculate the periodogram (power magnitude of FFT) of each frame.
4. Apply the MEL filters to each periodogram.
5. Apply log function to the output of MEL filters.
6. Apply the Discrete Cosine Transformation (type 2) to keep most relevant coefficients (2-13). The other coefficients are discarded since they represent rapid changes which are irrelevant for speech recognition.

7. De-emphasize higher frequency Mel coefficients using sinusoidal liftering (filtering in MEL domain) with a parameter of 22.

After calculating the MFCCs for a DUS segment, a feature matrix of dimension 12 by F , where F is the total frames of a segment, was obtained. The final MFCC coefficients for a segment were calculated using the strategy described by Zaidan and Salam (2016) [409], where six features are extracted for each coefficient. That is, the min, max, mean, median, mode, and variance of each coefficient in each frame. As a result, a total of 72 (6×12) features are obtained for each DUS segment.

4.3.6 Feature selection

Given that the feature vector contained 88 features per segment, a feature selection technique was used to reduce redundancy, classification error and computation time. The maximum Relevance Minimum Redundancy (mRMR) algorithm was chosen for this purpose. This algorithm was performed using a two-stage algorithm proposed in [285], which first uses filter techniques based on maximal relevance D and minimal redundancy R to find the compact set of features, calculated as:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c) \quad (4.3)$$

$$\min R(S, c), D = \frac{1}{|S^2|} \sum_{x_i, x_j \in S} I(x_i, x_j), \quad (4.4)$$

where S is the feature set, x_i and x_j are individual features, I is the mutual information and c is the target class. Optimizing for maximal relevance D and minimal redundancy R simultaneously, $mRMR$ is denoted as:

$$\max \phi(D, R), \phi = D - R. \quad (4.5)$$

Once the compact set was found, a wrapping technique was applied to further reduce the number of features. Specifically, backtracking selection was performed, thereby iteratively excluding the less significant features of the compact set until improvement was steady.

4.3.7 Classification

A two-step approach was developed to classify the DUS quality. The first step aimed to remove silent segments of data by using only variance as feature, which should always be a very small value for silent segments. A binary logistic regression classifier was trained using silent vs non-silent segments (all classes except silent).

The second step involved the use of a Support Vector Machine (SVM), which was shown to be reliable for classifying between good and poor classes recorded with the same device used in earlier related work [373, 351]. In this new work, the SVM was multiclass (four classes: *Good Quality*, *Poor Quality*, *Interference* and *Talking*) using the one-vs-one approach, which creates a total of $\frac{n(n-1)}{2}$ classifiers, classifying a sample based on class vote counting. This approach was implemented using the LibSVM suite [66]. A Gaussian radial basis function was chosen for the kernel which is parameterized by σ . Together with the soft margin hyperparameter (C), it was optimized using a grid search and five-fold cross-validation on the training set. The grid search was defined by $C \in \{2^{-3}, 2^{-1}, \dots, 2^5\}$ and $\sigma \in \{2^{-5}, 2^{-2}, \dots, 2^2\}$.

The test and training sets were normalized by subtracting the mean of the respective feature vector and dividing it by its standard deviation computed in the training data only.

4.3.8 Performance assessment

To assess the classification system presented here, the dataset was split into two equal subsets; the training set contained 107 recordings from 73 subjects and test set

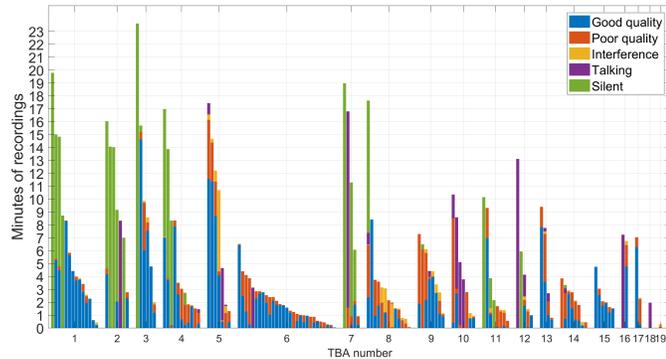


Figure 4.6: Durations of recordings taken from each subject (pregnant mother) in the data set. Subjects are grouped by the TBA who made the recordings ordered by the quantity of data recorded. For each TBA, subjects are ranked by total quantity of recording time.

contained 88 recordings belonging to the remaining 73 subjects. To avoid any bias between the training and test set, the subjects were randomly distributed, and then the rank sum test was performed to ensure the absence of any statistically significant differences in blood pressure, heart rate, gestational age, and TBAs' number of years of experience, between subjects in the two sets.

For the nested five-fold cross validation, subjects were stratified - i.e. distributed to one out of five possible folds, thus ensuring that features belonging to a given subject were present only in one fold. However, since the number of classes per subject was different in the training set (See Figure 4.6), the data set was class-unbalanced. To address this problem, a similar strategy as that presented in Valderrama *et al.* (2017) [373] was used; the number of classes for each fold was balanced using sampling with replacement (bootstrapping). For each fold, 10 samples per class per feature were randomly selected from the fold subjects' samples using sampling with replacement. The selection was performed in proportion to the relative contribution of each subject's data in each fold. Using the balanced folds, the pair of parameters C and σ with the highest recall was selected. Using this parameter pair, the final model was trained using all the balanced training data.

To assess the performance of the classifier on the test set, the logistic regression

model (for silent vs. non-silent classification) was first applied. Those samples classified as non-silent were then input to the multiclass SVM. A confusion matrix was calculated for the final result. This process was repeated 100 times, randomly selecting different samples in the sampling with replacement procedure. Likewise, this process was repeated for each of the possible window length (0.75, 1.50, 2.25, 3.00 and 3.75 s). In addition to the confusion matrix, global recall, precision, and the F1 score were calculated using micro-averaging and macro-averaging for the individual recall and precision of each class. Micro-averaging pools per-document decisions across classes, and then computes an effectiveness measure on the pooled contingency table. Macro-averaging computes a simple average over classes - the difference between the two can be large.

Micro-averaged precision (μPr), recall (μRe) and the geometric average of the two metrics, micro-averaged F1 ($\mu F1$), were calculated as:

$$\mu Pr = \frac{\sum_{i=1}^5 TP_i}{\sum_{i=1}^5 TP_i + \sum_{i=1}^5 FP_i}, \quad (4.6)$$

$$\mu Re = \frac{\sum_{i=1}^5 TP_i}{\sum_{i=1}^5 TP_i + \sum_{i=1}^5 FN_i}, \quad (4.7)$$

$$\mu F1 = 2 \frac{Pr_{micro} \times Re_{micro}}{Pr_{micro} + Re_{micro}}, \quad (4.8)$$

where i is the index of each of the five classes (*Good Quality*, *Poor Quality*, *Interference*, *Silent* and *Talking*); TP is the true positive of the $i - th$ class, FP is the false positive of the $i - th$ class; and FN is the false negative of the $i - th$ class.

Similarly, the macro-averaged Pr, Re and F1 (MPr , MRe , and $MF1$) metrics were calculated as:

$$MPr = \frac{1}{5} \sum_{i=1}^5 \frac{TP_i}{TP_i + FP_i}, \quad (4.9)$$

$$MRe = \frac{1}{5} \sum_{i=1}^5 \frac{TP_i}{TP_i + FN_i}, \quad (4.10)$$

$$MF1 = 2 \frac{Pr_{macro} \times Re_{macro}}{Pre_{macro} + Re_{macro}}. \quad (4.11)$$

After calculating the performance of the classifier, the same performance evaluation procedure was repeated for the best window length using only the features obtained with the mRMR feature selection algorithm.

4.3.9 Processing time

A final experiment was performed taking into account the processing time for feature extraction. Specifically, features were removed from the final model in order of the slowest to fastest processing time, and then the classifier was retrained and tested at each step to evaluate the performance of any ensemble of feature which cumulatively can be run under a second on a standard processor (an Intel(R) Xeon(R) CPU E5-2660 v2 @2.20GHz, on which we report all results here).

As a final step, mRMR was again applied to the real time features to determine if a small subset of them could perform almost as well as all the real time features. This was to ensure real time performance on even a weak smartphone processor. In our field work, we use a 1 GHz dual-core ARM Cortex-A9 running on Android 4.0, which is considerably slower [270]. To estimate the ratio of computational speed between the two processors (and hence make a reasonable estimate about whether the algorithms presented here could run in real time on the field device), we evaluated a standard FFT library [108] on both devices. An arbitrary 3.75 s DUS segment was presented to the FFT library 100 times, and the compute time was recorded and averaged.

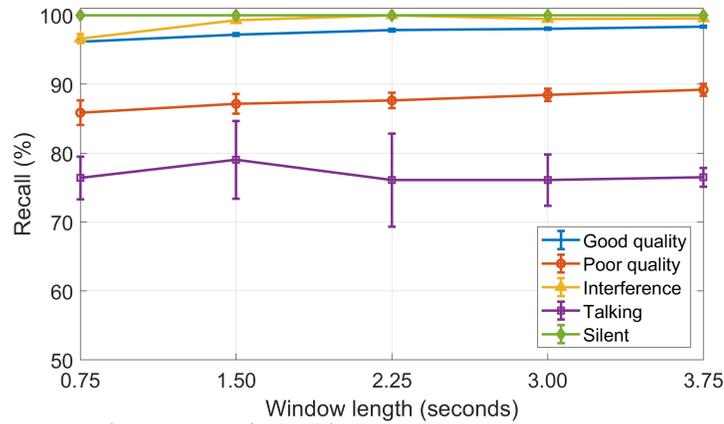


Figure 4.7: Error bars (median +/- IQR) per-class recall as a function of window length over 100 iterations of balanced bootstrap resampling of the unbalanced data.

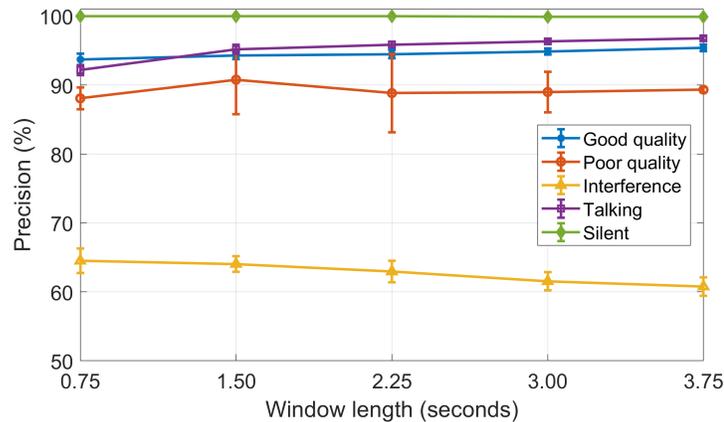


Figure 4.8: Error bars (median +/- IQR) per-class precision as a function of window length over the 100 iterations.

4.4 Results

4.4.1 Model performance

Figure 4.7 and Figure 4.8 respectively illustrate the per-class recall and precision as a function of window length. Precision was acceptable for all window sizes for all classes except ‘interference’. Recall was acceptable for all classes except ‘talking’. In general, a window size of between 1.5 s and 2.25 s gave best results, although there was not a large window size effect. However, it is notable that the IQR decreased as the window length increased.

Table 4.3: Confusion matrix of the classifier assessed on test data for window length of 3.75 s, and repeated 100 iterations. At each iteration, the data are sampled randomly with replacement, balancing the classes of the training set.

		Actual					Total
		<i>Good Quality</i>	<i>Poor Quality</i>	<i>Interference</i>	<i>Talking</i>	<i>Silent</i>	
Predicted	<i>Good Quality</i>	932,613	13,226	1,505	901	0	948,245
	<i>Poor Quality</i>	36,144	383,326	3,308	5,942	297	429,017
	<i>Interference</i>	0	0	42,899	189	0	43,088
	<i>Talking</i>	9,043	32,848	22,888	212,168	750	277,697
	<i>Silent</i>	0	0	0	0	951,053	951,053
Total		977,800	429,400	70,600	219,200	952,100	2,649,100

Table 4.4: Median global metrics of classifier on DUS segments (assessed on test data), based on 100 iterations. Recall, precision, and F1 score were aggregated for all classes using both micro-average and macro-average approaches.

Metric (%)	Window length (second)				
	0.75	1.50	2.25	3.00	3.75
μ Re	93.9	95.3	95.5	95.9	96.3
μ Pr	93.0	94.3	94.4	94.8	95.2
μ F1	93.5	94.8	94.9	95.3	95.7
<i>M</i> Re	87.5	88.7	88.6	88.5	88.4
<i>M</i> Pr	91.0	92.6	92.3	92.4	92.8
<i>M</i> F1	89.3	90.7	90.4	90.4	90.5

Table 4.3 presents the confusion matrix of the resultant classifier on the test data for window length of 3.75 s. (Confusion matrices of other window lengths are omitted for brevity). Correct classification of the majority of ‘talking’ segments was achieved, but a considerable number of other classes were also labeled as ‘talking’. Specifically, approximately 32% of interference segments were classified as talking, thus reducing the recall of this latter class. Similarly, precision of talking is affected because of the large number of interference segments classified as talking.

To provide an overall metric of the classifier presented here, a global precision, recall, and F1 score were calculated using both micro-average (eqs. 4.6, 4.7, 4.8) and macro-average approaches (eqs. 4.9, 4.10, 4.11). As can be seen from Table 4.4, both approaches (micro and macro) provide a global F1 score higher than 90%. Also, for both approaches, global recall, precision, and F1 score increased as window length increases, with only a marginal effect on window size.

Table 4.5: Cumulative performance when adding individual features selected by the mRMR algorithm ranked in order of execution time on a Intel(R) Xeon(R) CPU E5-2660 v2 @2.20GHz processor for a 3.75 segment of DUS with individual computation times. The final column indicates the performance of the classifier using the feature and all features above it in the table. (Note that due to rounding, the cumulative execution time is not exactly the sum of the individual times.)

Feature	Computation (& Cumulative) time (ms)	$\mu F1$	$MF1$
Wavelet energy percentage level 1 (1000-2000 Hz)	0.03 (0.03)	68.2	56.5
Wavelet energy percentage level 2 (500-1000 Hz)	0.01 (0.04)	77.0	60.5
MFCCs using max function coeff 5, 11, 12, 13	8 (8)	80.2	68.0
MFCCs using mean function coeff 6, 11	0.05 (8)	87.2	79.5
MFCCs using min function coeff 3,13	0.05 (8)	91.9	86.7
MFCCs using median function coeff 2, 8	0.27 (8)	93.3	88.7
MFCCs using std function coeff 9, 11, 12, 13	0.05 (8)	93.5	88.8
PSD ratio 4 (800 Hz)	11 (19)	95.6	93.1
PSD ratio 2 (217 Hz)	0.12 (19)	95.6	92.9
PSD ratio 3 (434 Hz)	0.13 (20)	95.8	93.9
<i>F0</i> (fundamental frequency)	21 (40)	95.9	93.9
<i>SQI 2</i> (linear resampling autocorrelation)	1028 (1068)	96.3	94.3
<i>SQI 3</i> (<i>DTW</i>)	3061 (4129)	96.5	94.4
<i>SQI 4</i> (<i>wDTW</i>)	4851 (8980)	96.8	94.5

Table 4.6: Median performance obtained on the held out test data by the retrained classifier using the selected features by the mRMR algorithm for a window length of 3.75 s, based on 100 bootstrap iterations.

Metric	<i>Good Quality</i>	<i>Poor Quality</i>	<i>Interference</i>	<i>Talking</i>	<i>Silent</i>
Re	96.6	92.8	94.8	89.5	99.9
Pr	97.3	89.2	88.7	95.8	100.0

4.4.2 Feature selection and real time performance

Table 4.5 provides the 23 features selected by the mRMR algorithm and ranked based in the median computation time for extracting them from a DUS segment of 3.75 s. Cumulative compute time and held out F1 score are also provided. Those features which prevent the real-time execution of the routine are italicized. Figure 4.9 illustrates the F1 global scores obtained as a function of number of features (backward selection, removing the slowest feature each time) and compute time. Table 4.6 summarizes test set performance per class on the reduced feature vectors (for real time operation). Note that recall and precision for *Talking* and *Interference* improved.

After applying the mRMR algorithm to the 83 features corresponding to features with executions time lower than 100 ms, a total of 17 features were selected (see Table 4.7 and Figure 4.10). Using this reduced features set, a $\mu F1$ score of 96.0 and

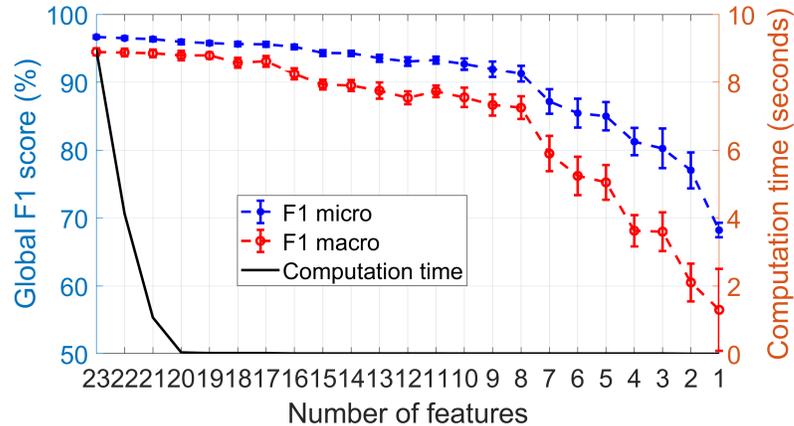


Figure 4.9: Median F1 scores (\pm IQR) of μ F1 (dashed blue line) and MF1 (dashed red line) for the different combinations of the mRMR-selected features. Black solid line represents the corresponding median computation time required for extracting the features from a DUS segment of 3.5 s. All results are for 100 bootstrap iterations.

Table 4.7: Optimize feature vector selected by the mRMR algorithm applied to those features whose median computation time for extraction from a 3.75 second segment is less than 100 ms. The features are ranked based on the individual execution time. τ is the cumulative time to compute the features in the corresponding row and all the features in the preceding rows. G, P, I, T, and S represent *Good Quality*, *Poor Quality*, *Interference*, *Talking* and *Silent* segments of data respectively.

Feature	τ (ms)	μ F1	MF1	Recall					Precision				
				G	P	I	T	S	G	N	I	T	S
Wavelet % energy at (1000-2000 Hz)	0.03	74.5	58.4	89.7	11.2	51.1	16.0	99.9	87.8	50.9	10.5	25.0	100.0
Wavelet % energy percentage at (500-1000 Hz)	0.03	80.7	63.7	85.0	54.1	52.8	31.0	99.9	89.0	62.0	16.9	46.1	100.0
Wavelet % energy percentage at (250-500 Hz)	0.04	82.9	65.2	89.8	60.2	45.3	33.5	99.9	90.7	69.4	16.7	47.5	100.0
MFCCs using max function coeff 5, 12, 13	7.91	84.0	67.4	90.0	60.3	45.0	43.7	99.9	91.9	67.0	17.7	58.2	100.0
MFCCs using mean function coeff 6	7.93	91.4	83.4	96.3	80.1	86.4	63.0	99.9	93.8	86.5	48.8	80.1	100.0
MFCCs using median function coeff 2, 8	8.20	94.4	89.4	95.9	86.1	89.2	82.3	99.9	95.7	87.3	68.7	89.4	100.0
MFCCs using std function coeff 9, 11, 12, 13	8.25	95.0	90.6	96.6	87.7	90.1	83.7	99.9	96.2	87.7	72.3	92.7	100.0
PSD ratio 4 (800 Hz)	19.20	95.6	93.2	96.5	89.5	95.8	87.3	99.9	96.1	88.2	85.5	94.8	100.0
PSD ratio 2 (217 Hz)	19.30	95.7	93.5	96.4	89.8	95.3	87.9	99.9	96.2	87.6	87.3	95.1	100.0
PSD ratio 3 (434 Hz)	19.46	95.8	94.2	96.3	90.2	94.7	88.4	99.9	96.2	86.8	96.1	94.8	100.0
F0 (fundamental frequency)	40.01	96.0	94.5	96.4	90.6	97.2	88.9	99.9	96.2	88.6	92.8	95.9	100.0

MF1 score of 94.5 were obtained, which are comparable to that using the 23 features selected using mRMR on all 88 features. Computational times for extracting features to a DUS segment of 3.75 seconds was reduced to 40 ms from 8980 ms. Moreover, recall and precision for each remained greater than or equal than 88.9% using the selected 17 features. We note that no individual feature gave a μ F1 score of 75.4 and MF1 score of 62.1, demonstrating that the multivariate model provided a significant improvement of any individual features.

The FFT benchmark test resulted into an averaged time of 1.88 ms for the pro-

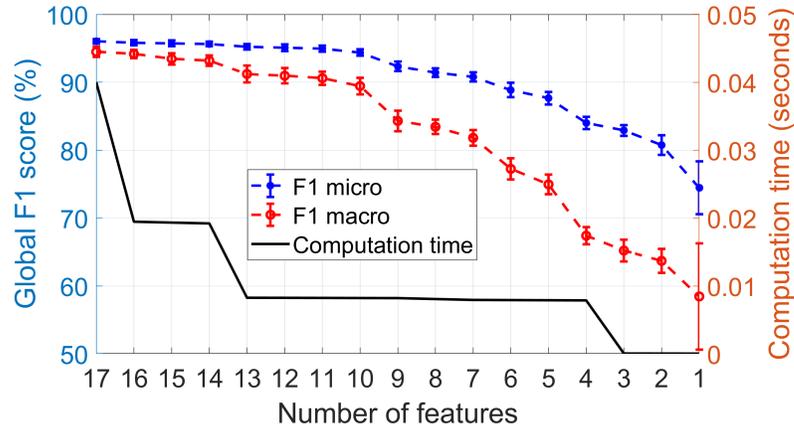


Figure 4.10: Median μ F1 (dashed blue line) and MF1 (dashed red line) scores (\pm IQR) for increasing numbers of features in the classifier. Only 85 features were considered, with individual execution time of less than 100 ms, and only 17 features were selected using mRMR. Black solid line represents the corresponding median computation time required for extracting the features from a DUS segment of 3.5 s. All results represent averages over 100 bootstrap iterations.

cessor on the Samsung S3 mini and 0.21 ms for the faster server processor on which all results presented here were derived. This suggests that the smartphone processor will compute metrics nine times slower than the times reported here. In other words, the time needed for extracting the 17 features presented above for a 3.75 seconds is approximately 360 ms. However, we note that some of the features other than the FFT-based ones, may lead to small differences in timing, since the chip architectures are significantly different in terms of cache and threading. Moreover, other demands from the operating system may lead to variance in compute time.

4.5 Discussion

The results presented here indicate that is possible to accurately classify different types of DUS recordings quality by their etiology in real-time, and thereby provide an instantaneous feedback system to healthcare workers using a DUS device. As for every application, the trade-off between keeping a low quality recording or discarding a good quality recording is always the issue. In this work, these issues were balanced;

however, it may make more sense to adjust the operating point through time.

Among the DUS quality categories, acceptable classification performance for *Good Quality* data was achieved for the real time classifier (Pr= 96.2% and Re=96.4%), ensuring that only a low percentage of good segments are discarded, thereby reducing the unnecessary retaking of the recordings and creating annoyance for the end user. Even though the number of classes in this work was increased from two to five, compared to previous works [373, 353, 351], high performances were seen across categories, with $\mu F1=96.0$ and $MF1=94.5$.

While the results on optimizing the temporal window length indicate that there is no clear window size that provides an optimal classification, in our work the 3.75 second window appears to present no large disadvantage. This length agrees with that defined for the Dawes/Redman approach [85] for computerized fetal analysis. Furthermore, a 3.75 second window length exhibited a lower IQR for all classes.

Reducing the number features from 88 to 17, greatly reduced computation time from almost 10 seconds to 40 ms on the training platform (equating to 360 ms on the low cost field smartphone, , which will allow the implementation of a real time quality feedback system). The feature reduction also increased the recall for classifying *Talking* segments and the precision for *Interference* segments. This indicates that the values of some of the original 88 features for the *Talking* and *Interference* classes overlapped significantly, and that the overlap was reduced after feature selection. Although when using the 17 feature real-time version of the classifier, the lowest Pr for *Talking* and Re for *Interference* segments compared to other segments, the improvement after feature selection is notable. The high real-time F1 scores indicate the potential of the method presented here, which is essential for informing the TBA of the reason for low DUS quality, thus providing a simple method for the TBAs to remediate the problem at source.

Finally we note that the features chosen are unlikely to be dependent on geographic

location, because they are general to human speech, but not language specific. However, if used in a different non-rural context, retraining may be required for noises such as electromechanical interference from surgical equipment, industrial ambient noise masking speech, or other industrial noise pollutants such as car alarms. Nevertheless, these are likely to be misclassified as the wrong noise type, rather than as good quality data.

4.6 Conclusions

This work presents an approach to assessing DUS signal quality captured by TBAs with minimal training in the use of the equipment (a low-cost Doppler device connected to a smartphone). By extracting various signal features over short windows and using a cascaded classification approach (a binary logistic regression and a multiclass SVM), this article demonstrates that it is possible to effectively differentiate between high quality 1D-DUS recordings and the key noises encountered in the challenging environment of the Guatemalan rural setting. The method presented here would be portable back to our smartphone app used by TBAs, thereby providing them with a tool to identify sources of error during recording. Ensuring quality in the recordings captured by TBAs should facilitate a reliable fetal monitoring system in LMIC settings for use by non-experts and potentially contribute to reducing the burden of perinatal health issues.

Chapter 5

Fetal heart estimation from 1D-DUS signal

5.1 Abstract

Objective: Open research on fetal heart rate (FHR) estimation is relatively rare, and evidence for the utility of metrics derived from Doppler ultrasound devices has historically remained hidden in the proprietary documentation of commercial entities, thereby inhibiting its assessment and improvement. Nevertheless, recent studies have attempted to improve FHR estimation; however, these methods were developed and tested using datasets composed of few subjects and are therefore unlikely to be generalizable on a population level. The work presented here introduces a reproducible and generalizable autocorrelation (AC)-based method for FHR estimation from one-dimensional Doppler ultrasound (1D-DUS) signals.

Approach: Simultaneous fetal electrocardiogram (fECG) and 1D-DUS signals generated by a hand-held Doppler transducer in a fixed position were captured by trained healthcare workers in a European hospital. The fECG QRS complexes were identified using a previously published fECG extraction algorithm and were then over-read to

ensure accuracy. An AC-based method to estimate FHR was then developed on this data, using a total of 721 1D-DUS segments, each 3.75 s long, and parameters were tuned with Bayesian optimization. The trained FHR estimator was tested on two additional (independent) hand-annotated Doppler-only datasets recorded with the same device but on different populations: one composed of 3938 segments (from 99 fetuses) acquired in rural Guatemala, and another composed of 894 segments (from 17 fetuses) recorded in a hospital in the UK.

Main results: The proposed AC-based method was able to estimate FHR within 10% of the reference FHR values 96% of the time, with an accuracy of 97% for manually identified good quality segments in both of the independent test sets.

Significance: This is the first work to publish open source code for FHR estimation from 1D-DUS data. The method was shown to satisfy estimations within 10% of the reference FHR values and it therefore defines a minimum accuracy for the field to match or surpass. Our work establishes a basis from which future methods can be developed to more accurately estimate FHR variability for assessing fetal well-being from 1D-DUS signals.

5.2 Introduction

During pregnancy, fetal cardiac monitoring is a common method for identifying fetal abnormalities in the second and third gestational trimesters [318]. This identification process is performed by examining fetal heart rate (FHR) variations in signals between 10 to 60 minutes, using epochs of 3.75 s as is described in the Dawes/Redman system [279, 85]. Based on the observable variations, physicians and midwives attempt to identify abnormal patterns indicative of stress or adverse outcomes, in theory facilitating timely intervention if required [280, 246, 142].

The most common method to monitor FHR and uterine contractions is using

Doppler ultrasound devices. These devices use a transducer to measure the change in frequency of a reflected acoustic wave from an object moving relative to the acoustic source. In the case of FHR monitoring, the change of frequency is caused by cardiac wall and valve movements, and sometimes blood flow. These physical movements are recorded as a one-dimensional ultrasound (1D-DUS) signal, which is demodulated into an audio recording so that the operator of the device can hear the resultant heart beat changes in the range of human hearing. This ‘sound’ can then be analyzed to estimate FHR, much like any other cardiac signal. To achieve this, Doppler ultrasound devices use autocorrelation (AC)-based approaches, in an attempt to identify the dominant frequency in a given band (or delay window).

Although Doppler ultrasound devices have been widely used for decades to estimate FHR, one of their limitations is that commercial monitoring companies have not disclosed the details of the algorithms such that ‘someone skilled in the art’ could actually reproduce the exact approach [241]. Therefore, the accuracy of Doppler-based FHR estimation systems remain unknown and replication or improvement of such approaches is entirely inhibited.

Despite the difficulty of accessing the FHR estimation source code, relatively little research has been published concerning improving Doppler-based FHR estimation [173, 175, 288, 308]. The majority of publications on this topic that do exist were aimed at improving FHR estimations by comparing the proposed method to estimations provided by direct scalp fECG taken from the fetal head during labor. Although these methods used an accepted reference methods with which to compare their results, they had some limitations. Specifically, they developed and tested their methods using only one dataset composed of only a few patients (between 1 to 15) [173, 175, 288, 308], optimizing parameters for their specific dataset without holding out data for validation, making it highly unlikely that the developed algorithms would have any generalizability beyond the small number of individuals studied [214].

In addition to AC-based methods, in recent years, researchers have also proposed a different approach to estimate FHR from 1D-DUS signals. Specifically, Al-Angari *et al.* (2017) proposed the use of empirical mode decomposition (EMD) of the 1D-DUS signals and the kurtosis of the instantaneous mode function (IMF) as a measure of FHR [6]. This method was trained using abdominal fECG as a reference. The method was developed using a total of 44 1D-DUS signals of one-minute length each individually extracted from healthy single pregnant women within 24 and 42 gestational weeks. The authors presented evidence that the EMD-kurtosis method achieved higher accuracy than AC-based methods, specially for low SNR signals. However, the parameters of this method, such as the number of IMFs and the window size to calculate kurtosis, were optimized using all of the dataset, again leading to likely overfitting on the limited data used.

Therefore, the aim of this work is to develop a reproducible and generalizable AC-based method able to accurately estimate FHR from 1D-DUS taken with an inexpensive hand-held transducer [351, 354] from subjects of different populations. Although AC-based methods are affected by the inherent smoothing or averaging of the autocorrelation (ACF) [176, 201, 65], an AC-based FHR estimator remains a computationally inexpensive approach and, if it is confirmed to be accurate to estimate FHR, it can be a valuable first indicator for fetal well-being in itself. Moreover, a recent work has shown that an accurate AC-based estimator can be the basis of more advanced analysis to segment 1D-DUS into beat-to-beat segments [351], from which FHR variability analysis can be performed.

To ensure an accurate FHR estimation, the method was optimized by comparing to FHR obtained from a simultaneously recorded abdominal fECG as a benchmark. After the method was optimized, it was tested on two independent datasets and different populations, and over the second and third trimesters in order to test any dependency on gestational age.

5.3 Methods

5.3.1 Databases

This work used three different 1D-DUS databases collected over a period of four years. All signals were acquired using a hand-held 1D-DUS 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, captured using a Samsung S4, S3 mini or S4 mini and stored as uncompressed WAV files at 7056 Kbps (16 bits).

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 [23]. The database included data from 16 volunteers with pregnancies between the 20th and 27th week of gestation, including pathological cases such as Intrauterine 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 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).

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 [353, 351, 373].

Guatemala RCT Database

This dataset was collected as part of a randomized controlled 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.

In a recent study on 1D-DUS signal quality assessment we found that quality is critical for the estimation of fetal heart rate [374]. Since processing low quality data would lead to spurious comparisons with incorrect heart rates, only 1D-DUS signals

that had been manually annotated as “good quality” were used. In total, there were 195 1D-DUS signals recorded from 146 pregnant women, who were carrying singletons between the fifth and ninth month of gestation.

5.3.2 Manual Heart Rate Estimation

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, and in the case of the Leipzig University Hospital database, on the simultaneous fECG windows as well.

Annotation of the Leipzig University Hospital(LUH) Database

For the LUH 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 extended Kalman smoother was used [38]. Then, the filtered fECG and the 1D-DUS signal were resampled to 4 Khz, and were displayed in a graphical user-interface (GUI) (Figure 5.1), 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 [279, 85]. Furthermore, in our previous work this window length was shown to be suitable for assessing 1D-DUS quality acquired with the same hand-held device used in this study [374]. 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 FECSYN toolbox [38]. These estimated fECG QRS peak times were taken as guide for locating the beats in the 1D-DUS signal.

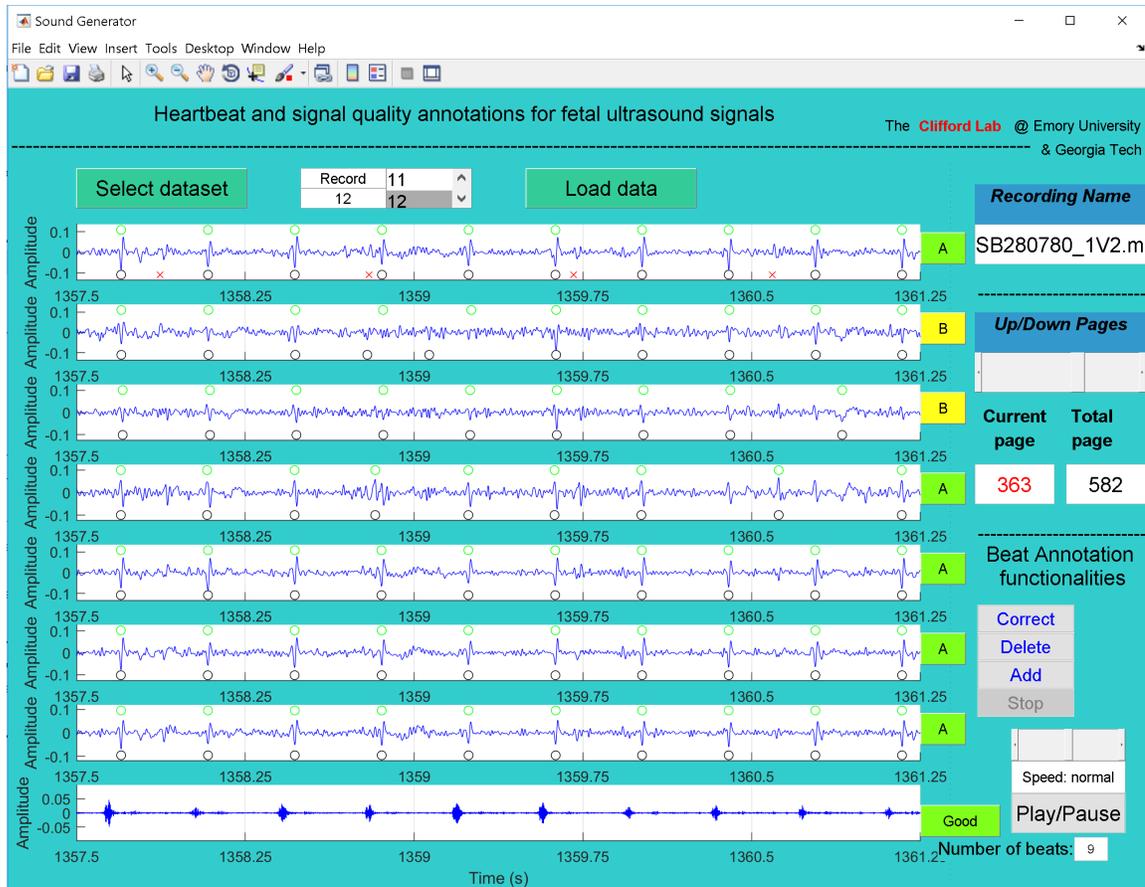


Figure 5.1: GUI used to annotate LUH database. The first seven channels correspond to abdominal fECG after filtering maternal components using an extended Kalman smoother and peak detection of fetal (black circles) and maternal beats (red crosses, upper plot) [38]. The last channel in the plot is the 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).

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. 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 [374], namely, good, poor, electrical interference, talking, silent, or unsure. Since 1D-DUS quality may affect the FHR estimation [351], 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 chan-

nel)

- 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 5.1), 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 5.2 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.

The reciprocal of the median of the interval between fECG peak times in a 3.75 s segment (scaled by 60) was used to estimate the reference FHR. The median is highly robust to missing or extra peaks [73].

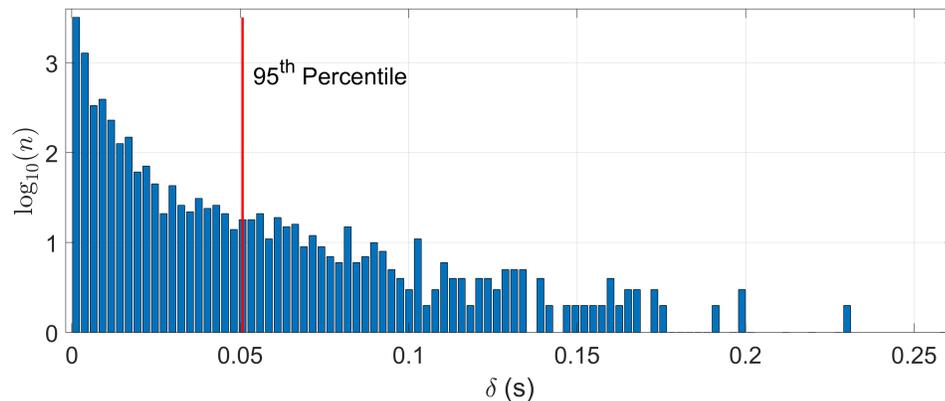


Figure 5.2: Histogram of time differences between beat time locations from two independent annotators. The horizontal axis, δ , is the difference in seconds of the two annotations for the fECG peak timing. The vertical axis represents a logarithmic scale of the count (n) of each difference. Note that 95% of the annotations differed by at most 0.05 s.

Annotation of Oxford JR Database

Each of the 1 minute-length 1D-DUS signals were labeled by three different expert annotators using a Matlab GUI. Each reviewer independently labeled 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 ± 15 BPM from the last corrected peak. More description of the method can be found in [351].

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

Annotations of Guatemala RCT Database

For this dataset, only segments that had been manually classified as good quality in our earlier work [374] were used. Since in Doppler ultrasound each cardiac cycle is represented by a combination of cardiac wall and valve movements [327], 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 reactions, keyboard delays, etc., add in large variable time delays [354]. To address this problem, we designed a Matlab GUI (Figure 5.3) 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.

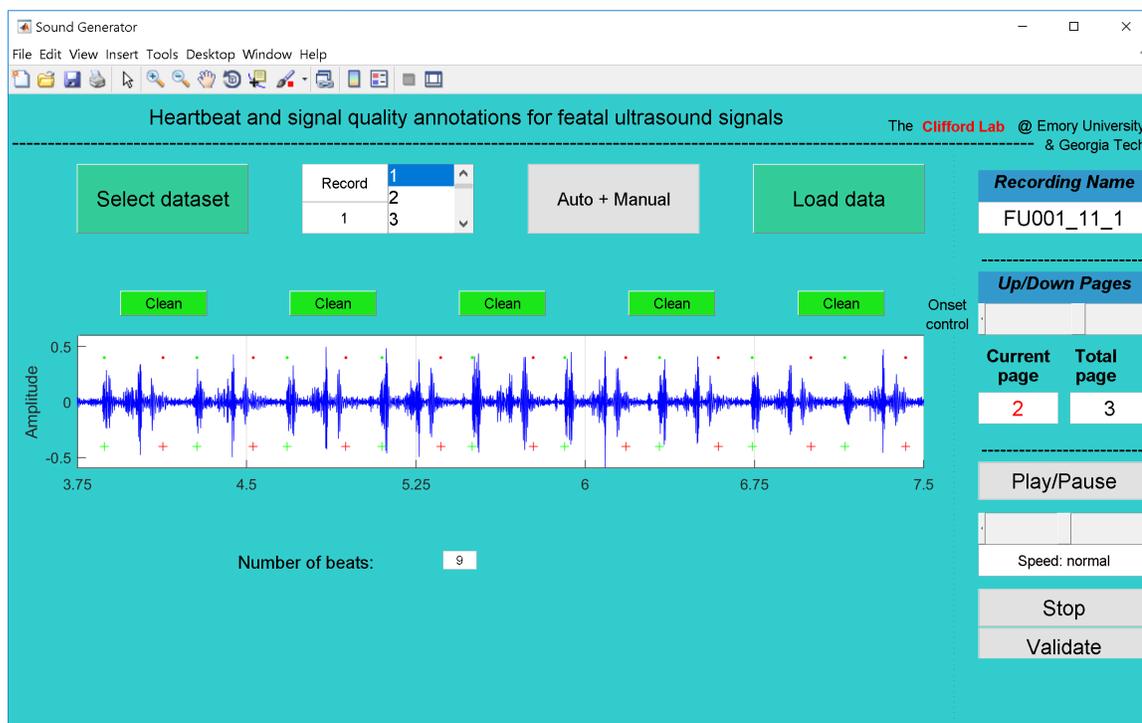


Figure 5.3: GUI used to manually annotate the number of beats in the Guatemala RCT dataset. Annotators listened to each 3.75 s segment, counting and recording the number of audible beats.

Table 5.1: Summary of final datasets used for developing fetal heart rate estimator. 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 remaining two dataset were used as independent test sets. For the Oxford dataset, the number of good and excellent quality segments are shown in parenthesis.

Database	Number of subjects	Number of 3.75 s segment	GA range (weeks)	Additional recordings	Use
Leipzig (LUH)	5	721	20-27	fECG, maternal ECG	Training
Oxford JR	17	894 (482)	20-38	1D-DUS quality labels	Testing
Guatemala RCT	99	3938	20-40	blood pressure, maternal HR	Testing

5.3.3 Datasets

The number of 1D-DUS segments obtained for each dataset after the manual FHR estimation process are displayed in Table 5.1.

Figure 5.4 shows the distribution of the manual FHR estimations of the three datasets used in this work. Note that they form very different distributions, having been derived from datasets with different gestational ages. The LUH dataset contained subjects with lower gestational ages (generally less than 27 weeks) than the Oxford and Guatemala RCT datasets and FHR is known to be higher earlier at earlier stages of pregnancy [166]. Accordingly, the histogram of the Leipzig manual

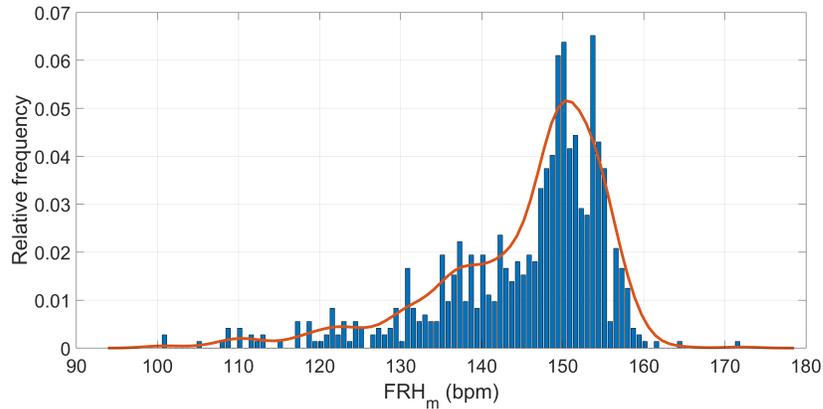
FHR estimations had a peak around 150 BPM, whereas the other two datasets had a peak around 135 BPM. Note also that the majority of FHR estimated values are found concentrated into in the 110 to 160 BPM normal interval [291]. The LUH and Oxford JR datasets appear to have continuous distributions since manual annotations were performed based on the timing of the peak. On the other hand, the Guatemala RCT dataset contained a much more pronounced quantization because the annotation of this dataset was based on the number of audible beats in a 3.75 s window, ranging between 112 to 192 BPM (7 to 12 beats in 3.75 s).

5.3.4 Heart rate estimator

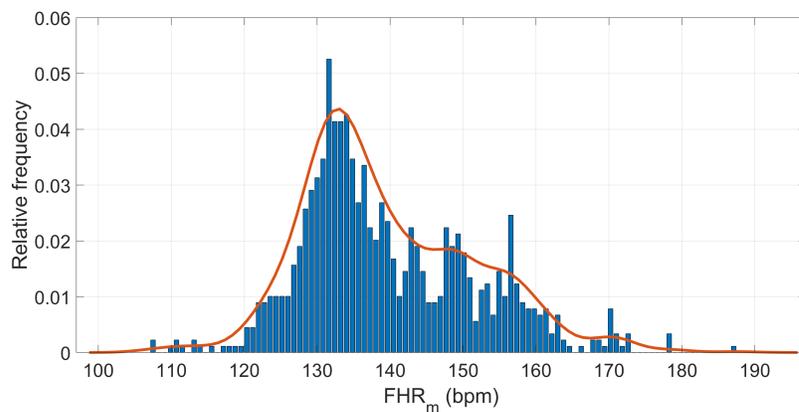
The following subsections describe our AC-based method to estimate FHR from 1D-DUS signals. The corresponding source code is available from a public repository [376] under a BSD Clause 2 license.

Noise removal

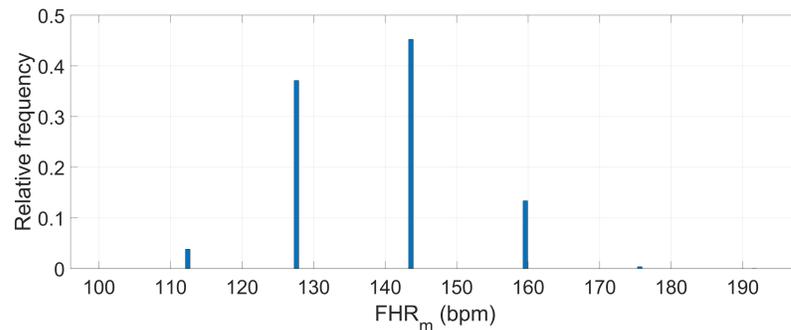
As extreme noise affects the estimator prediction, noise spikes were removed, using the algorithm presented in [321]. This algorithm splits the 1D-DUS segment into windows of a specified interval. In this work, the interval was defined to be 0.75 s to contain at least one beat based on a heart rate of 80 BPM, which is the lower bound on a normality interval [291]. For each of the windows, the maximum absolute amplitude (MAA) was calculated. Then, windows whose MAA value was higher than a threshold, defined as three times the median MAA of the segments, were marked as containing spikes. To remove spikes, the marked window with the greatest MAA was selected, defining the spike interval as the first zero-crossing before and after the MAA. That interval was set to zero. After this, the median windows MAA and threshold were recalculated, and the procedure was repeated until there were not windows with MAA greater than the threshold.



(a)



(b)



(c)

Figure 5.4: Manual FHR (FHR_m) estimations for datasets: (a) Leipzig dataset; (b) Oxford dataset; (c) Guatemala RCT dataset. The solid (red) lines represent the probability distributions smoothed using a normal kernel function [53]. The Leipzig and Oxford datasets appear to have continuous distributions, while the Guatemala RCT contained a much more pronounced quantization.

Frequency range of interest

The cardiac frequency range for the device used here was estimated from observations in the current literature. For instance, Tutschek *et al.* (2003)[365] found that from

the 12th gestational week, the peak cardiac wall velocities could be measured by tissue Doppler echocardiography. This finding suggests that cardiac wall movement, in particular ventricular motion, may be present in any acquired 1D-DUS signal. According to the authors, the axial cardiac wall velocities of the right and left ventricle are described by the following second order data fitting models:

$$\begin{aligned} V_{RV} &= 0.017x^2 + 0.5944x + 9.0522, \\ V_{LV} &= 0.009x^2 + 0.2104x + 5.0742, \end{aligned} \tag{5.1}$$

where V_{RV} and V_{LV} , in cm/s, are the axial cardiac wall velocities of the right and left ventricle, respectively; and x is the gestational age in weeks.

To define the cardiac frequency range, the empirical models of the cardiac wall velocities (Eq. 5.1) were combined with the equation of Doppler magnitude frequency shift f_D , defined as [189]:

$$f_D = \frac{2f_o}{c} V \cos \theta, \tag{5.2}$$

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 angle the between ultrasound beam and the surface in radians.

The hand-held Doppler ultrasound device used in this work has an f_o of 3.3 MHz. The speed of sound in human tissue (c) in standard-compliant ultrasound machines is 1540 (m/s) [104]. The maximum f_D is achieved when θ is 0 *rad*, whereas it is minimum when θ is $\pi/2$ *rad*. Using a gestational range interval from 20 to 40 weeks, the maximum and minimum values for measured change in frequency (Hz) were estimated (see Table 5.2). Based on Table 5.2, the cardiac frequency range was extracted using a 25-600 Hz bandpass filter. Low frequencies (below 25 Hz) were

removed to reduce disturbances introduced by fetal or device movement.

Table 5.2: Expected minimum and maximum values for measured change in frequency (Hz) from over a gestational age ranging from 20 to 40 weeks. RV and LV stands for right and left ventricular, respectively.

Variables	RV	LV	RV	LV
GA [weeks]	20		40	
V [m/s]	0.04	0.04	0.12	0.11
f_{Dmin} [Hz]	0	0	0	0
f_{Dmax} [Hz]	169.89	191.41	534.69	473.92

Homomorphic envelope

Given the capacity of complex homomorphic filtering to extract envelopes from physiological processes [302], such as electromyographic or respiratory signals, in this work we applied this method to the 1D-DUS segments. The rationale behind this technique is to assume that physiological processes are generated by a multiplicative system as:

$$s(t) = A_m(t) \cos(w_c t), \quad (5.3)$$

where $A_m(t)$ is the modulating signal carrying the information, and $\cos(w_c t)$ is the carrier signal oscillating at a frequency of w_c . However, since the carrier frequency, w_c , is unknown, the demodulation is not straightforward. It can however, be achieved by transforming the modulation process into a system of linear filters. Although this transformation can be performed as:

$$\log(s(t)) = \log(A_m(t)) + \log(\cos(w_c t)), \quad (5.4)$$

applying it directly to real signals can cause a major concern since logarithm has a singularity at zero. Nevertheless, this problem is alleviated by using the analytical signals as is presented in [302]. Specifically, the Hilbert transform is used, which

converts the real function, $s(t)$, into an analytical function as:

$$z(t) = s(t) + i\tilde{s}(t), \quad (5.5)$$

where $\tilde{s}(t)$ is the Hilbert transform of $s(t)$. The analytical function, $z(t)$, has the advantage that its magnitude, $|z(t)|$, is equivalent to that of the message signal in the amplitude modulation process. Thus, the magnitude, $|z(t)|$, is decomposed into amplitude and oscillating components by using Eq. 5.4. Then, the amplitude components are kept by using a low-pass filter. Finally, taking the exponential of the low-pass filter output, the homomorphic envelope of $s(t)$ is extracted.

Automatic Fetal Heart Rate Estimation

The FHR was estimated by applying the ACF to the homomorphic envelope of the 1D-DUS signal. The ACF was estimated in time-domain using a rectangular window from lag 0 to the length of the 1D-DUs segment (i.e. $3.75 \times 4000 = 15000$ samples). As described in Box *et al.* (2015)[54], the $k - th$ lag correlation, r_k , of a sequence $y[n]$ was calculated as:

$$r_k = \frac{c_k}{c_0}, \quad (5.6)$$

where c_0 is the sample variance of $y[n]$, and c_k is:

$$c_k = \frac{1}{T} \sum_{i=0}^{T-k} (y[i] - \bar{y})(y[i+k] - \bar{y}), \quad (5.7)$$

where \bar{y} is the mean of $y[n]$, and T the length of $y[n]$.

The envelope periodicity was determined from the ACF using the algorithm shown in Figure 5.5 and in Appendix B. This algorithm found peaks within a window range of possible FHR values, which was defined by a lower interval value within 0.25-0.3 s (200-240 BPM) to an upper interval value within 0.8-1.0 s (60-75 BPM). Since the

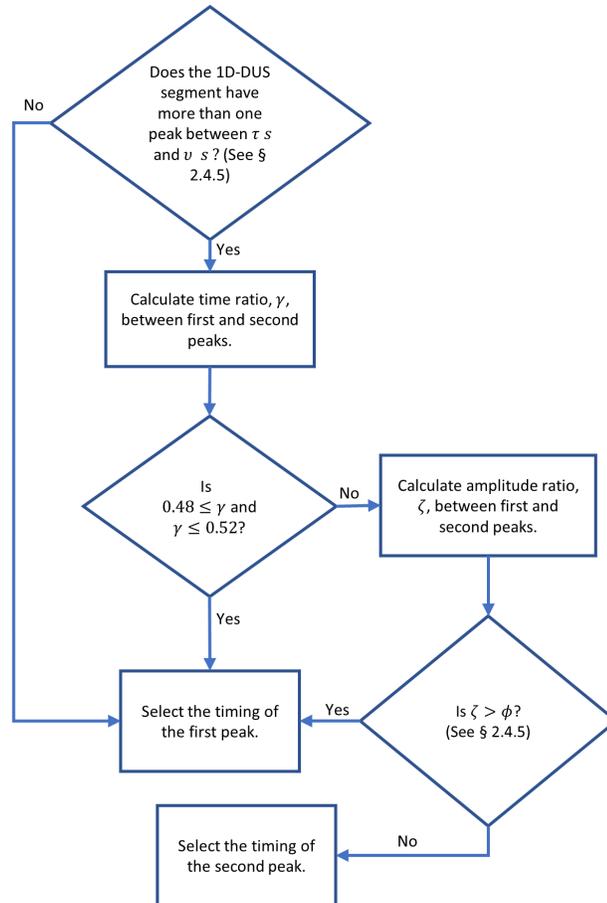


Figure 5.5: Flow diagram of algorithm used for finding the peak in the ACF to determine the heart rate period. Pseudocode for this flow diagram is provided in Appendix B.

maximum window length using these values could be 0.75 s, the window may include at most two prominent peaks in the ACF.

of the 1D-DUS segments in the training data contained harmonics within this region. In case the peaks are harmonics, the temporal value (lag) of the first peak is selected as that which represents the periodicity. Otherwise, the first peak is selected only if the amplitude ratio between the peaks, ζ , is greater than a threshold, which was determined by parameter optimization described in the following section.

For instance, Figure 5.6 shows an example for one of the 1D-DUS segments. As can be seen, there are two peaks inside the search window. In this case, if the highest amplitude peak were selected without comparison to the lower amplitude peak, it would lead to an incorrect estimate of the heart rate period (i.e a period of 0.749 s or

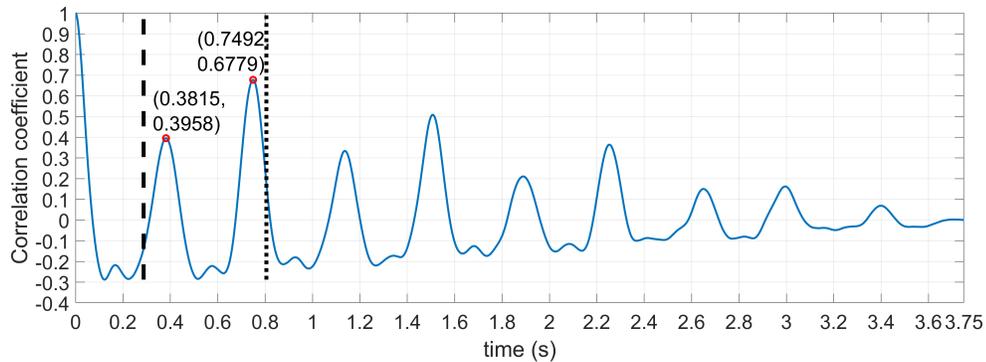


Figure 5.6: ACF for a 3.75 1D-DUS segment. Vertical dashed and dotted lines indicates the respective start and end points of the window within which a search is made for the peak related to the heart rate period. Peaks inside the window search are compared using algorithm showed in Figure 5.5 to find the right time location for the signal period.

80.1 BPM). The first peak correctly corresponds to a period of 0.382 s (157.3 BPM). The ratio of the peaks, γ , is $0.382/0.749$, which is close to 0.5, indicating the second peak is a harmonic and should be ignored.

Once the algorithm selects a peak time, the FHR of the 1D-DUS segment (i.e. periodicity) is determined as $FHR = 60/l$ BPM, where l is the time location in second of the selected peak.

Parameter optimization

There are four parameters that must be optimized to estimate the FHR. The first is the cut-off frequency of the low-pass filter used to extract the envelope from the complex signal. The second (τ) and third (ν) parameters are the lower and upper bounds of the interval of the window used for finding peaks in the autocorrelation. The fourth parameter, ϕ , is the threshold used for comparing the peaks amplitude within the search window.

The four FHR estimator parameters were determined using Bayesian optimization with 1000 iterations over the training set (a subset of the Leipzig dataset). The parameter search space was defined as shown in Table 5.3. The Bayesian optimization used the mean square error (MSE) of the difference between the manual FHR anno-

tation and the FHR estimations of the training subjects (MSE_{m-t}) as the objective function to be optimized. In detail, given a parameter tuple within a search space bounds, the objective function first calculated the MSE_{m-t} of each training subject, and then averaged the individual MSE_{m-t} values.

At the beginning, the Bayesian optimization algorithm defined a Gaussian process (GP) for the objective function with mean 0 and covariance kernel defined by the Automatic Relevance Determination (ARD) Matérn 5/2 kernel [341]. At each iteration, the GP was updated using the posterior probability of the new evaluated tuple. The tuple to evaluate at each iteration was selected using the expected-improvement acquisition function (EI) [128]. This function selected the tuple that maximized the expected improvement on the objective function based on the associated GP. The parameter *expectation ratio* of the acquisition function was set to 0.5 to avoid being stuck at local maxima, thus giving the same weight to explore and exploit parameter tuples in the solution space. After the 1000 iterations, the parameter tuple with the minimum value of the objective function was selected as the parameters for the FHR estimator.

Table 5.3: Parameter intervals used in the Bayesian optimization process.

Parameter	Lower bound	Upper bound
Cut-off frequency (Hz)	2	15
Minimum Period (s)	0.25	0.3
Maximum Period (s)	0.8	1.0
Threshold	0.55	0.99

5.3.5 Performance assessment

Since the LHU dataset contained both 1D-DUS signals and simultaneous abdominal fECG, it provided the most accurate reference point from which to start. We trained our FHR estimator using 60% of this dataset so that the fECG annotations were used as the primary reference. The remaining 40% of the LHU dataset was used as validation to reduce the chances of model overfit. The trained model was then tested

with the Oxford JR and the Guatemala RCT datasets to evaluate generalizability performance. It is important to note that these two datasets were excluded from the training procedure because these latter two datasets contained no fetal ECG as a reference and including all of the populations in the training data would lead to overfitting [214]. This process is somewhat similar to that of the commercial device setting, in which algorithms are trained on relatively few recordings, but are expected to work on a large variety of populations.

Training process

The Leipzig dataset was stratified by subject into training and validation sets, ensuring that manual FHR values between the two groups did not differ significantly (p-value > 0.05 , Wilcoxon rank sum test). Thus, the training set contained a total of 430 3.75 s segments from three subjects, while the validation set contained 291 segments of 3.75 s from the remaining two subjects.

After optimizing parameters on the 1D-DUS segments of the training subjects, the FHR estimator was tested on the segments of the validation subjects. As well as ensuring a good out of sample optimization, the validation set was also used to check that the method did not estimate the maternal heart rate (MHR) rather than FHR. To this end, the automatic FHR estimations were compared to the maternal heart rate to ensure that the method was indeed estimating the fetal cardiac cycle. (The MHR was calculated using the maternal ECG simultaneously recorded with the 1D-DUS signals).

Testing process

Using the optimal parameters obtained from the Leipzig dataset, the fetal heart rate estimator was tested on the Oxford and Guatemala RCT datasets separately to see if the algorithm generalized to multiple different populations. For both test

datasets, the automatic FHR estimations were compared to those manually estimated by the annotators. To perform this comparison two methods were used. First, we defined a positive percentage of agreement (PPA) as a FHR estimation that are within 10% of a manual FHR. This bound was deviated from previous works, which have performed comparison to asses the equivalence in success rate, reliability and accuracy between FHR measurements obtained from ultrasound CTG and abdominal fetal ECG [299, 300]. This PPA has been also used to compare ultrasound CTG and abdominal fetal ECG fetal scalp ECG [105, 78]. Second, to show the bias and the level of agreement of our FHR estimator, Bland-Altman plots [49] were created.

In the Oxford dataset, FHR estimation was also calculated according to quality classes in the dataset. For the Guatemala RCT dataset, additional stratification by gestational age (GA) was performed to check if this variable was a confounder for FHR estimation. GA comparison was performed comparing all possible pairs of the available months (5 to 9 months). The significance level ($\alpha = 0.05$) of the hypothesis test performed on the 10 possible pairs, $\binom{5}{2}$, was corrected using the Bonferroni technique ($\alpha/10 = 0.005$).

Comparison between Oxford JR and Guatemala RCT datasets

As datasets have different distributions (see Figure 5.4), we performed a two-sided Wilcoxon rank sum test to determine if the median of the errors between the Oxford JR and the Guatemala RCT datasets were significantly different. (A one sample Kolmogorov-Smirnov test was first applied to confirm that the error distributions were not normally distributed.) Only good quality 1D-DUS segments of both datasets were used so that meaningless comparisons between spurious estimates were not made. A total of 3934 and 479 segments for the Guatemala and Oxford datasets, respectively, were available for analysis.

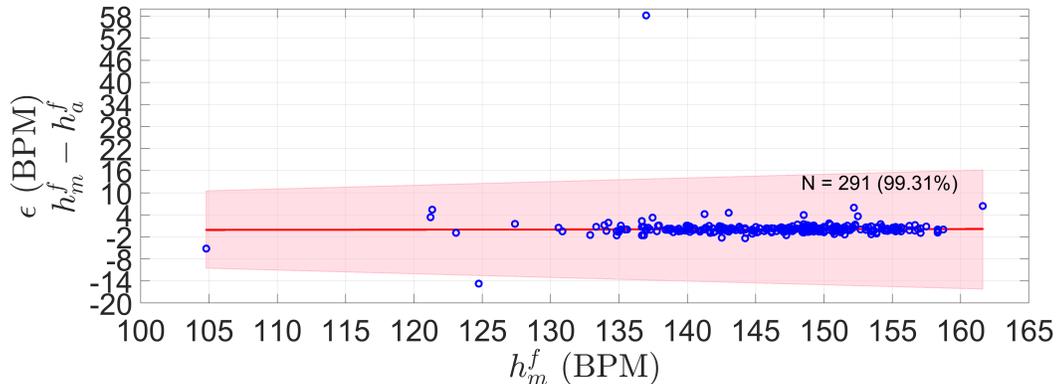


Figure 5.7: Error (ϵ) (difference) between manual FHR (h_m^f) and automatic FHR (h_a^f) of the LHU validation set. Shaded area shows the 10% PPA bounds (error no greater than 10 percent of the h_m^f). N stands for the number of segments in the validation set, and in parenthesis is the percentage of segment satisfying the PPA. The red line is the robust least square fit of ϵ along the manual FHR estimations ($\epsilon = -0.571 + 0.04h_m^f; r^2 = 0.99$).

5.4 Results

5.4.1 Optimized parameters and validation stage

Table 5.4 shows the best parameters obtained through the Bayesian optimization process. Window search was found to range between 0.29 s to 0.84 s, which was equivalent to 71.5 and 208.8 BPM, respectively.

Table 5.4: Best tuple of parameters found by the Bayesian optimization process.

Cut-off frequency (Hz)	Minimum Period (s)	Maximum Period (s)	Threshold
14.8	0.287	0.839	0.650

Figures 5.7 and 5.8 show the performance of the trained FHR estimator on the validation set. As can be seen, 99% of the segments were within the PPA bounds, whereas the difference between the manual (h_m^f) and automatic FHR (h_a^f) was close to zero with a RSME of 3.68 BPM. The low bias suggests that optimized parameters were not overfitted, and therefore the resulting algorithm may be appropriate to other unseen data.

In addition, to further evaluate the performance on the validation set, the automatic FHR was also compared to MHR (Figure 5.9). Only 1 out of 291 estimations was in the maternal range, thus suggesting that the 430 Doppler segments used for

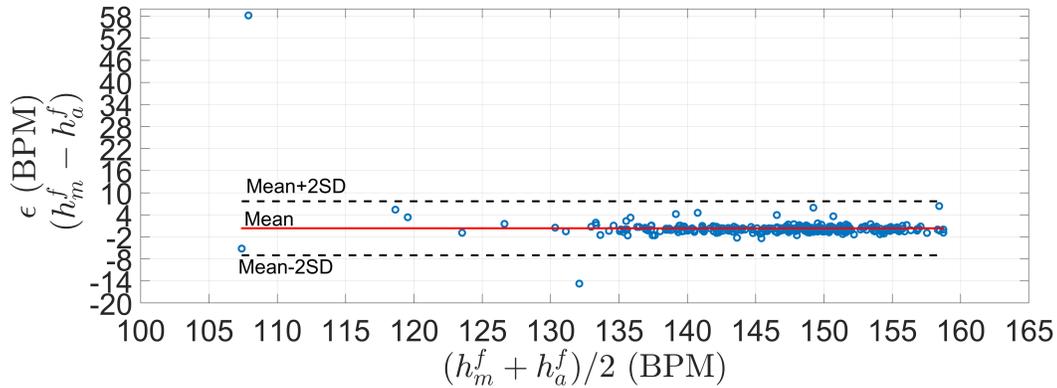


Figure 5.8: Bland-Altman plots for the LHU validation set. Mean (μ) and standard deviation (σ) of ϵ were 0.34 and 3.67 BPM, respectively. Limit of agreements ($\mu + 2\sigma$) was $[-7, 7.68]$ BPM. RSME was 3.68 BPM

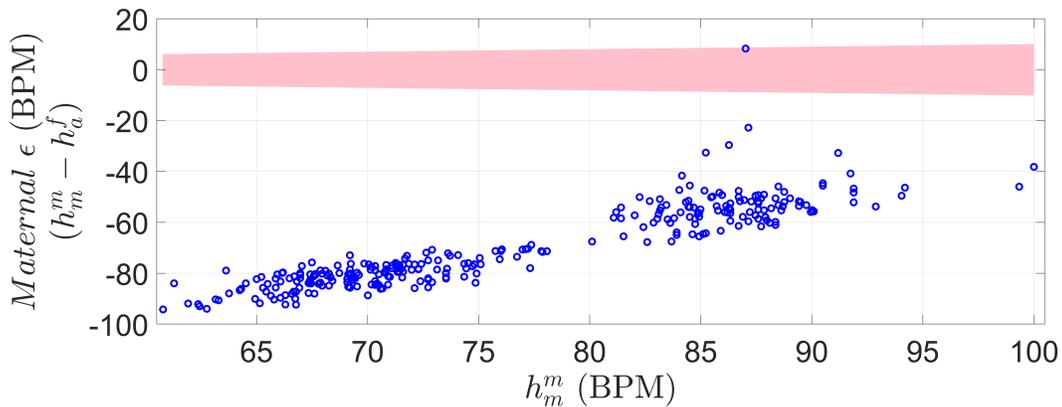


Figure 5.9: Maternal-Fetal difference in heart rate, for the validation set. This quantity was defined as the difference between maternal heart rate (h_m^m) and automatic FHR (h_a^f) estimation. The shaded area represents bounds in which automatic HR estimation is considered maternal instead of fetal based on the PPA (h_a^f estimation within 10% of the recorded h_m^m). Only one estimation out of the 291 was within the shaded area, indicating a possible maternal recording (from maternal arteries).

training the estimator indeed corresponded to fetal rather than maternal cardiac activity.

5.4.2 Oxford dataset

Using the optimal parameters, the fetal heart rate estimator was applied to the 1D-DUS segments of the Oxford JR dataset, which included data of different quality levels. Figure 5.10 shows the difference between the manual FHR and the automatic FHR discriminated by quality class. For all the signal quality classes, the FHR

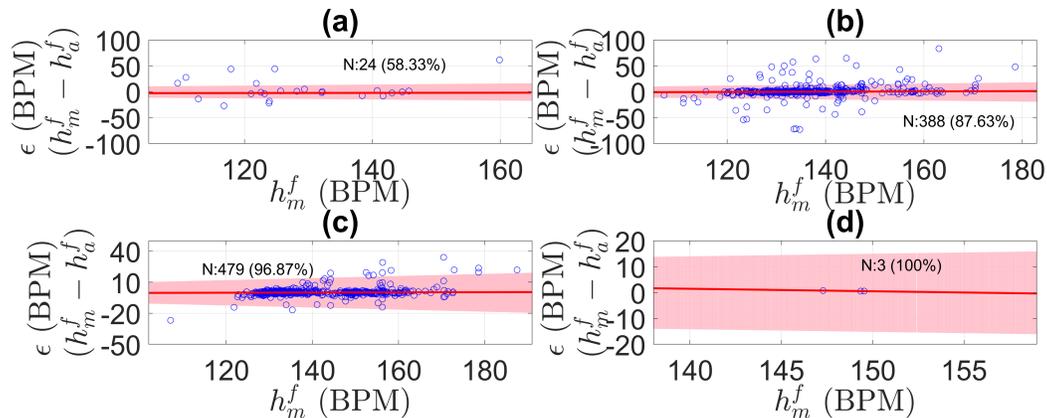


Figure 5.10: Error (ϵ) in fetal heart rate estimation discriminated by signal quality for the Oxford dataset: (a) poor quality, (b) intermediate quality, (c) good quality, (d) excellent quality. ϵ was defined as the difference between manual FHR (h_m^f) and automatic FHR (h_a^f) estimations. Shaded area shows the PPA error bounds (error no greater than 10 percent of the h_m^f). N stands for the number of segments in that signal quality class, and in parenthesis is the percentage of segment within the tolerance error bounds. The red line is the robust least square fit of ϵ along the manual FHR estimations (r^2 was 0.995, 0.997, 0.997, and 1.0 for poor, intermediate, good, and excellent quality, respectively).

estimation algorithm satisfied the PPA bounds over 87% of the time for data of intermediate or better quality.

Likewise, Table 5.5 shows the bias and variance of the difference between manual and automatic FHR estimations. As 1D-DUS quality increased, the bias and the variance decreased, thereby suggesting that signal quality category is critical for FHR estimation from 1D-DUS signals.

Table 5.5: Mean and standard deviation, level of agreement, and RSME of the Bland-Altman analysis for the Oxford dataset discriminated by signal quality class.

Quality class	Mean (BPM)	Standard deviation (BPM)	Level of agreement (BPM)	RSME (BPM)
Poor	4.72	21.28	[-37.84, 47.28]	21.36
Intermediate	1.08	14.31	[-27.54, 29.69]	14.33
Good	0.52	4.62	[-8.72, 9.76]	4.65
Excellent	0.62	0.12	[0.38, 0.86]	0.63

5.4.3 Guatemala RCT dataset

Figure 5.12 shows the estimator performance for all the Guatemala RCT segments. As can be seen, 97% of the automatic FHR estimations fit the PPA bounds. Since the difference between two adjacent values of h_m^f is 16 BPM, ϵ increased in 4.48 BPM

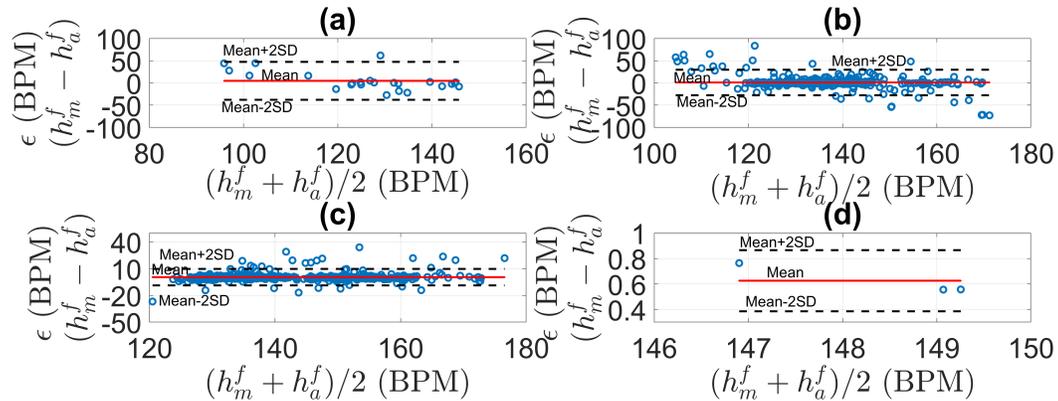


Figure 5.11: Bland-Altman plots for the fetal heart rate estimation discriminated by signal quality for the Oxford dataset: (a) poor quality, (b) intermediate quality, (c) good quality, (d) excellent quality. Mean and standard deviation, level of agreement, and RSME for each class are shown in Table 5.5.

(0.28×16) as h_m^f goes from one available value to the next one.

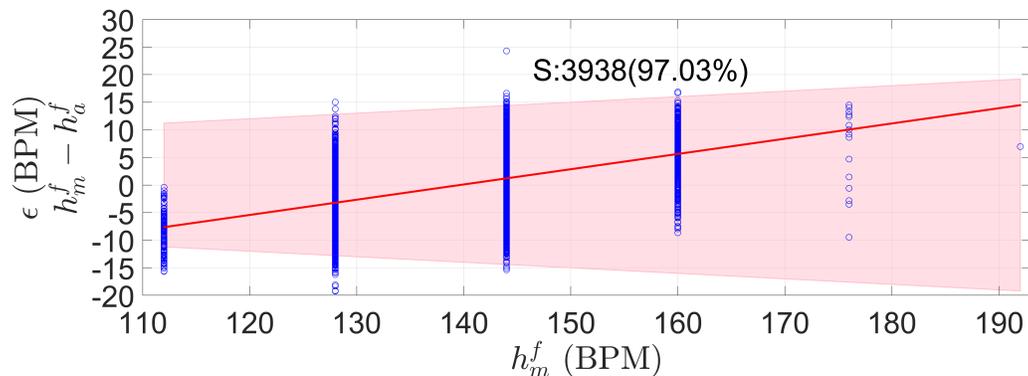


Figure 5.12: Error (ϵ) measured as the difference between manual (h_m^f) and automatic FHR (h_a^f) estimations for Guatemala RCT dataset. The shaded area represents the PPA bound, defined as no greater than 10 percent of the h_m^f . S stands for the number of segments, and in parenthesis is the percentage of segment satisfying the PPA bounds. Red line is the robust least square of ϵ along the manual FHR estimations ($\epsilon = -38.63 + 0.28h_m^f; r^2 = 0.9962$).

Figure 5.13 shows a Bland-Altman analysis of the manual and automatic estimates of FHR in the Guatemalan RCT database. The mean difference between FHR estimates was close to zero, whereas the RSME was 6.64 BPM. The low bias and variance indicate that the FHR estimator can also achieve accurate estimations for datasets with a completely different distribution than that of the training set (see Figure 5.4).

Figure 5.14 shows the results for the different GA found in the dataset. As can

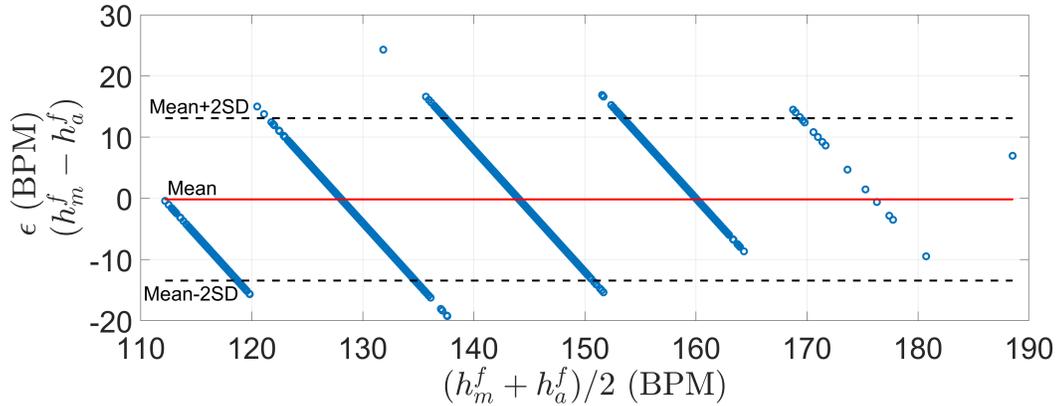


Figure 5.13: Bland-Altman plots for the Guatemala RCT dataset. Mean (μ) and standard deviation (σ) of ϵ were -0.17 and 6.63 BPM, respectively. Limit of agreements ($\mu + 2\sigma$) was [-13.44, 13.10] BPM. RSME was 6.64 BPM

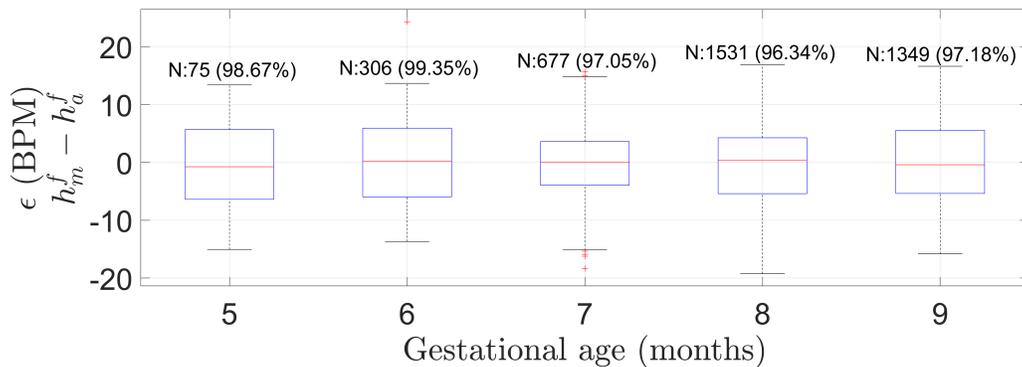


Figure 5.14: Box plot of the error (ϵ), defined as the difference between manual heart rate (h_m^f) and automatic FHR (h_a^f), for Guatemalan RCT dataset grouped by gestational age in months. N indicates the number of segments in the GA group and in parenthesis is the percentage of segments satisfying the PPA 10% error tolerance. FHR estimations did not statistically significantly differ among any possible pair of gestational months (Bonferroni corrected p-value of 0.005 (0.05/10); two-sided Wilcoxon rank sum test).

be observed, FHR estimation errors did not differ significantly statistically between gestational months (Bonferroni corrected p-value 0.005 (0.05/10); Wilcoxon rank sum test).

5.4.4 Comparison between Oxford JR and Guatemala RCT datasets

The medians of the distributions of the errors between h_m^f and h_a^f were not statistically significant (p=0.44; two-sided Wilcoxon rank sum test). Therefore, it seems that the

distribution type did not affect the performance of the FHR estimator.

5.5 Discussion

5.5.1 Interpretations of Findings

The performance of the AC-based method presented here indicates that FHR can be reliably estimated from 1D-DUS signals, particularly when coupled with a signal quality metric. Specifically, our method were able to estimate FHR within 10% of the manual estimated FHR, achieving these estimations with a low bias and variance. It is notable that our method accurately estimates FHR using autocorrelation, which is an inexpensive technique requiring few computational steps and little memory, thus saving cost and execution time. Presented results suggest that our method is a simple option able to achieve comparable FHR values to those provided by fECG, which is considered a validated reference.

In comparison with previous methods which reported a correlation coefficient, r , of 0.977 [288], and 0.992 [175] between estimated FHR and fECG measurements, our presented AC-method obtained an r coefficient of 0.9226 and 0.8319 for the good quality segments of the two test datasets (Oxford and Guatemala RCT). Although our values were lower, it should be noted that the methods cannot be directly compared since they were not tested on the same dataset. In fact, the methodology used here to train and test the FHR estimator indicates that parameters found with Bayesian optimization seem to be generalizable to other datasets. In contrast, previous methods were tested with few patients [175] or with recordings of a single patient [288]. Regarding the recently proposed EMD-Kurtosis method that claims to give more accurate estimations than standard AC-based methods [6], our method did not optimize the parameters on data used for both training and testing, thus making generalizability much more likely.

Our AC-based method demonstrated its robustness to the acquisition environment. Although the Leipzig and Oxford datasets were recorded in hospital environments, performed by medical professionals, the method also achieved equivalently satisfactory results for the Guatemala RCT dataset, which was acquired in rural areas by non-medically trained users. Likewise, on the Guatemala dataset, the FHR estimator achieved similar results for pregnancies with GAs greater than 20 weeks (5 months), thus suggesting that gestational age is not a confounding variable for the estimator (and indicating its promise for future work on predicting GA and growth-related problems). Moreover, unlike previous research that used data at intrapartum stages and direct fECG for training these methods [173, 175, 288], here we showed that FHR estimation from 1D-DUS is also possible for antepartum stages. Finally, based on MHR comparisons performed for the validation set of the Leipzig dataset, it is evident that the method presented here estimates fetal heart rate using fetal cardiac activity rather than maternal one.

The optimized parameters trained over the 60% of the LHU subjects were able to extend to the other two datasets (Oxford JR and Guatemala RCT) regardless of the fact that their distributions of the fetal heart rates were different in the other datasets (see Figure 5.4). Indeed, there was no statistically significant difference estimations of FHR on the test datasets.

Although the AC-based method obtained accurate estimations for all tested datasets, it is important to note that since the Guatemala dataset is more quantized than the Oxford dataset, there is a larger quantization error for the former dataset. Indeed, since the difference between two possible reported values over 3.75 s windows was 16 BPM ($\Delta=16$), the method has a quantization error of ± 8 BPM ($\Delta/2=8$), whereas the Oxford dataset did not have such a coarse quantization error.

5.5.2 Study Limitations

As previous research has reported, signal quality has been shown to be a relevant issue for heart rate estimation using AC-based methods [351, 6]. In fact, for the Oxford dataset, in which segments of different quality were available, the accuracy of the estimation depended on the 1D-DUS quality. This low accuracy for poor quality signals is explained by the fact that the ACF fails to provide a reliable determination of the signal periodicity when it contains additional spikes caused by noise. Nevertheless, the AC-based FHR estimation method presented here has been shown to be an accurate method across multiple datasets, mapping closely to industrial guidelines, for data of intermediate or better quality. We do not expect any system to provide good heart rate estimation when data are heavily corrupted, and in fact it is more important at that point to not report heart rate, but to report that the data are non-analyzable. In our earlier works, [353, 351, 373, 374], we demonstrated accurate methods for separating low 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.

Although all the datasets used here were acquired with the same hand-held device introduced in [351, 354], all datasets were recorded and manually annotated by independent volunteers, thereby reducing or eliminating any dependency among them. Future work should evaluate the performance of the method presented here using 1D-DUS acquired with different devices; however, promising results obtained here suggest that the method would estimate suitable results with minimal adjustment. One of these adjustments could be the frequency ranges for filtering the cardiac wall velocities since the ultrasound frequency might be different for other devices, thus modifying values presented in Table 5.2.

One limitation of our implementation was the small number of simultaneously acquired Doppler and fECG recordings available for training. The dataset was col-

lected by a partner group as an ancillary study to a larger study, and we had no control over it [23]. As a consequence, the quality was variable, and some recordings were discarded as the quality of fECG was not high enough to manually identify fetal heartbeats. Nevertheless, the promising results obtained on the two independent datasets suggest the potential for the method introduced here to estimate FHR from 1D-DUS signals.

This study was not intended for assessment of FHR during labor and delivery, but rather for hand-held point of care devices. However, we also note that we see no theoretical reasons that the conclusions reached in this paper would not hold during labor and delivery outside of the period of contractions (where the data are well known to be very noisy).

Based on parameters shown at Table 5.4, the AC-based method introduced in this work could detect FHR values ranging from 71.5 to 208.8 BPM. The lower bound at 71.5 BPM could be a drawback for the detection of some bradycardia abnormalities, which are associated with lower FHR values. Nevertheless, the upper bound to detect bradycardia is different among the most common fetal monitoring guidelines [319]. In particular, the International Federation of Gynecology and Obstetrics defines fetal bradycardia as a heart rate under 80 BPM, whereas the American College of Obstetricians and Gynecologists and the United Kingdom National Institute for Health and Care Excellence use 100 BPM and 110 BPM, respectively. Therefore, since our method is able to detect FHR estimations between 71.5 and 110, some bradycardia cases could be detected. We also note that the data used here is not relevant for labor, which is out of scope and not relevant to this research. For labor and delivery, the ECG has been shown to be superior [316]. We are specifically interested in tracking heart rate during the second and third trimesters before labor. Finally we note, that by definition, our data is representative of the patient population and encompasses the majority of heart rates we would encounter.

5.5.3 Future Directions

The rationale for the current research was to present a baseline for future FHR estimation methods. This serves several future directions of research and device design. For instance, the AC-based FHR estimation method presented in this work could help with the assessment of the accuracy of commercial Doppler transducers, which appear to use similar, but black-box AC-based methods. This work could improve the quality of regulatory submission and provide consumers with a more objective insight into the performance of a DUS FHR device. Furthermore, per [351], the method described in this paper could serve as an auxiliary method for initializing more accurate FHR and FHR variability estimators.

5.6 Conclusion

This work presents a simple but promising method to estimate FHR from 1D-DUS signals acquired using a hand-held Doppler transducer, obtaining estimations within 10% of the reference FHR values. The presented method is generalizable, in contrast to other methods presented in literature, and robust to the recording environment and operator skill. Therefore, the described AC-based method is valuable in itself as an estimator of FHR in all sorts of applications, including outside of the controlled hospital environment. By ensuring accurate estimation of FHR, a basis for fetal cardiac monitoring is provided, establishing a basis from which future methods can be developed to estimate FHRV more accurately for assessing fetal well-being from 1D-DUS signals. We encourage benchmarked contributions to our source code, which is freely available from a public repository [376] under a open source (BSD Clause 2) license.

Chapter 6

Estimating birth weight from postnatal weights

6.1 Abstract

Objective: Low birth weight is one of the leading contributors to global perinatal deaths. Detecting this problem close to birth enables the initiation of early intervention, thus reducing the long-term impact on the fetus. However, in low-and middle-income countries, sometimes newborns are weighted days or months after birth, thus challenging the identification of low birth weight. This study aims to estimate birth weight from observed postnatal weights recorded in a Guatemala highland community.

Approach: With 918 newborns recorded in postpartum visits at a Guatemalan highland community, we fitted traditional infant weight models (Count's and Reeds models). The model that fitted the observed data best was selected based on typical newborn weight patterns reported in medical literature and previous longitudinal studies. Then, estimated birth weights were determined using the weight gain percentage derived from the fitted weight curve.

Main results: The best model for both genders was the Reeds2 model, with a mean square error of 0.30 Kg^2 and 0.23 Kg^2 for male and female newborns, respectively. The fitted weight curves exhibited similar behavior to those reported in the literature, with a maximum weight loss around three to five days after birth, and birth weight recovery, on average, by day ten. Moreover, the estimated birth weight was consistent with the 2015 Guatemalan National survey, no having a statistically significant difference between the estimated birth weight and the reported survey birth weights (two-sided Wilcoxon rank-sum test; $\alpha = 0.05$).

Significance: By estimating birth weight at an opportune time, several days after birth, it may be possible to identify low birth weight more accurately, thus providing timely treatment when is required.

6.2 Introduction

Guatemala suffers the highest perinatal morbidity and mortality rates in Latin America, particularly affecting Mayan indigenous women in highland rural areas [404]. This high burden is a result of barriers, such as economic status, language, and culture, that limit the access to professional medical assistance for performing routine perinatal screening and medical referral. Due to these barriers, pregnant Guatemalan indigenous women are usually attended by traditional birth attendants (TBA), who lack access to adequate medical equipment and sufficient training for the identification of abnormal fetal development.

Low birth weight (LBW) is one of the leading contributors to global perinatal death rates, being the second cause after premature birth [86, 361]. LBW could be a consequence of either preterm birth (< 37 weeks) or intrauterine growth restriction (IUGR). However, the former is more common in industrialized countries, whereas the latter commonly occur in low-and middle-income (LMICs) [295, 384], in which

around 60% of LBW newborns are due to IUGR [205]. In fact, IUGR has a prevalence of around 11% in LMICs [89].

In an attempt to address IUGR in Guatemala, we have developed a smartphone-mediated affordable perinatal screening system in rural highland Guatemala [354, 231, 232]. This monitoring system allows Traditional Birth Attendants (TBAs), with minimal training, to use a pictogram- and audio-guided mobile application to assess the maternal and fetal well-being during perinatal visits. In these visits, TBAs ask standard symptom questions, take maternal blood pressure, and record One-Dimensional Doppler Ultrasound (1D-DUS) with a low-cost transducer connected to the smartphone.

During postpartum visits, TBAs register the newborn weight, among other parameters. A birth weight lower than 2500 g is an indicator of LBW. Therefore, in countries such as Guatemala, in which recent studies reported an LBW prevalence between 13% [52] and 14.6% [254], measuring birth weight is an absolute necessity. Thus, LBW newborns can start early treatment to alleviate short-and long-adverse consequences [86].

Unfortunately, in the Guatemalan rural highland areas is not always possible to register a newborn weight within a few days after birth due to the difficulties in following-up on patients [366]. Indeed, in our clinical trial, monitoring over 1000 women over two years [231], some of the postpartum visits were performed days or even weeks after birth. Moreover, the natural drop in body weight in the first week or two after birth, followed by the restoration of birth weight, makes the use of recorded weight problematic. One solution to this challenge is to translate the weight recorded after birth into an estimated birth weight using standardized infant weight charts or curves. However, standard weight charts have been constructed using populations from industrialized countries and may be inaccurate for an LMIC or rural population, particularly of non-European descent [264].

The work presented here introduces an approach for estimating newborn birth weights in Guatemalan highland using observed weights recorded days and weeks after birth. To this end, previously reported weight curve models were optimized to fit the observed weights of postpartum visits by using metrics based on infant weight development. The fitted model allows for the characterization of the weight loss and gain behavior of the population used in this study, as well as estimating birth weight.

6.3 Background

6.3.1 Infant weight development

After birth, newborns lose weight within the first days due to physiological diuresis and low initial enteral intake [217, 67, 378]. This weight loss is universal regardless of the feeding method [115] or the initial birth weight [326]. However, the maximum percentage of weight loss is not a total agreement, ranging among 7% to 10% [98]. After the occurrence of the maximum loss, the weight steady increases during the first months of life (see Figure 6.1).

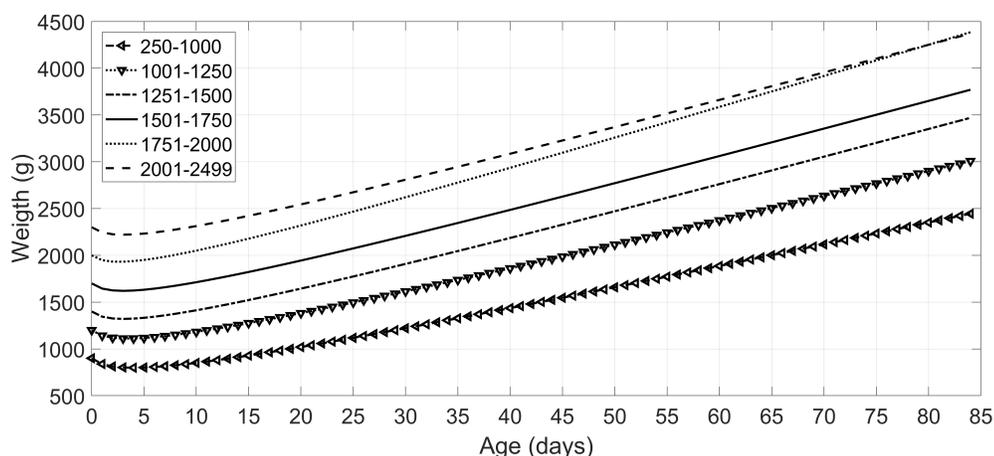


Figure 6.1: Expected infant weight curves as reported by [19], showing a maximum loss weight around the third to fifth days after birth. After this nadir, the weight steady increases during the first months of life.

In order to characterize infant weight patterns, different longitudinal studies have

been carried out among different countries. DiTomasso *et al.* (2018)[97] measured weight behavior of 134 newborns in a Rhode Island hospital, observing an average weight loss percentage of 8% around the third and fourth day, whereas the birth weight was recovered by the twelfth day. Likewise, Paul *et al.* (2016)[282] tracked the weight behavior of a large sample, 143,889 newborns, recorded in California. They found a range for percentage weight between 5.9% and 7.1%, occurring around the 61st and 65th hours after birth. After the loss weight nadir, the weight increased at a rate of 35 to 40 g/day, recovering the birth weight by day 10.

Although no study has been conducted for the Guatemalan population, some studies carried on Latin American LMICs have also reported similar findings than those previously mentioned [410, 80, 19, 276]. The reported average weight loss percentage for these studies was around 8% on the third or fourth day after birth. However, there was a variability for the recovery day and the weight increment rate, ranging between 5.7 to 19 days and between 13.9 to 40 g/day, respectively. All details of the previous infant weight development studies are shown in Table 6.1.

Table 6.1: Detail of infant weight development reported by previous published studies.

Study	Location	Sample size	Weight loss percentage (%)	Weight loss day	Recovery day	Weight rate (gr/day)
[97]	Rhode Island	134	[6-9]	[3-4]	[10-12]	39
[282]	California	143889	[5.9-7.1]	[2.5-2.8]	[9-10]	[35-40]
[410]	Mexico	101	[4.1-13.1]	[2.5-5.9]	[5.7-16.1]	[13.9-21.7]
[80]	Argentina	810	8	3	[8-10]	
[19]	Brazil	340	[5.9-9.7]	[4-5]	[16-19]	[22.8-35.99]
[276]	Uruguay	148	[8-9]	[2-3]	[8-10]	[30-40]

6.3.2 Weight curve models

Since the middle of the twentieth century, different non-linear models have been introduced for weight gain in infants and children [146]. The first introduced model was the Jenss [171], an exponential model, which showed suitable results for describing head-circumference and length growth, but not for describing the irregularity of weight patterns occurred in the first days after birth.

In addition to the Jenss model, three other non-linear models have been also used for fitting infant weight data. The first one was the Count's model (Eq. 6.1) proposed by Count (1942)[79] and modified by Shohoji and Sumiya (2001)[329]. The other two models, Reeds1 (Eq. 6.2) and Reeds2 (Eq. 6.3), were extensions of the Count's model developed by Berkey and Reed (1987) [46]. All these models were represented by several parameters, including the regression coefficients (a, b, c, d, e) ; age in months (t) ; and weight in Kg (y) .

$$y = a + bt + c \ln(t + 1) \quad (6.1)$$

$$y = a + bt + c \ln(t + 1) + \frac{d}{t + 1} \quad (6.2)$$

$$y = a + bt + c \ln(t + 1) + \frac{d}{t + 1} + \frac{e}{(t + 1)^2} \quad (6.3)$$

The weight curve models have been used in different studies to fit collected weight data. Within these studies, the Count's model has been the most commonly used. Thus, Peter *et al.* (2001)[286] fitted the Count's model for a sample of 1931 neonates collected between 1982 and 1990. Similarly, the Count's model was used in Brazil for describing the weight behavior of 340 newborns from birth to 12 weeks of life Anchieta *et al.* (2004)[19], as well as in Taiwan for examining the weight growth of infants exposed to some specific nutritional and ecological conditions [186].

All the models have also been compared to assess their fitting performance for weight data. Initially, Berkey (1982)[45] compared Jenss and Count's, reporting that Jenss performed better than Count's model as the latter challenged to fit data when it was spanned over six years. However, in another study, Peerson *et al.* (1993)[284] observed that any of these two models were able to fit weight recorded from 70 infants at the University of California Davis during their first 24 months. Nevertheless, in a more extensive study including five different models, Simondon *et al.* (1992)[335] found that the first-order Reed model was the best model for fitting weight data

collected from 95 infants from birth to 13 months of age in rural Congo.

In recent studies, quantile regression has been used over the traditional non-linear weight curve models. Indeed, the majority of weight models fitting after the 2010s applied non-linear quantile regression provided in modern programming languages [248, 282, 116, 393]. However, these models assume that the input is a longitudinal dataset, in which multiple samples are recorded for each patient along time.

6.3.3 Birth weight in Guatemala

Previous works have shown that the average birth weight in Guatemala is lower than the 3.5 Kg of European heritage newborns [168]. In fact, in a global maternity health registry study, Bose *et al.* (2015)[52] reported an average birth weight of 2,983.2 g (SD = 469.1 g) for a total of 30,262 deliveries. The total percentage of LBW newborns for that sample was 13%.

Similarly, in the Guatemalan national maternal and infant health survey [254], the averaged birth weight from 5,604 female newborns was 3,046.6 g (SD = 568.3 g), whereas for 6,071 male was 3,146.5 g (SD = 593.8 g). The percentage of LBW among rural and urban areas was similar, being 14.8% and 14.2%, respectively.

6.4 Methods

6.4.1 Database

Data used in this work was collected as a part of a perinatal care program conducted in rural highland Guatemala in the vicinity of Tecpan, Chimaltenango. This program was approved by the Institutional Review Boards of Emory University, the Wuqu' Kawoq | 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). In the

program, traditional birth attendants (TBA) were trained to use a mobile app to record perinatal information during regular visits. More details on the design and implementation of the data collection system, and the training of the TBAs can be found in [354] and [231, 232].

For this work, data recorded from 918 newborns in postpartum visits were used. For each newborn’s weight measurement was only performed once. These visits occurred on different postnatal days for each case, with an average at 31.36 days after birth (SD=32.86). In addition to measuring the weight, the TBAs also recorded the newborn’s birth date and gender. Moreover, in some cases, TBAs recorded the maternal age at birth, and the number of previous pregnancies (gravity). Table 6.2 shows available demographic information for the data used in this work.

As all data used in this work were recollected in the same rural community, pregnant women and newborns shared maternal nutrition, and delivery and feeding methods among them.

Table 6.2: Average demographics for the data used in this study. For each metric, the standard deviation and the number of samples available for that variable are shown in parenthesis.

Demographic variable	Total
Postnatal visit (days)	31.36 (SD= 32.86; N= 918)
Neonatal weight (Kg)	3.93 (SD= 2.73; N= 918)
Maternal age (years)	26.69 (SD= 8.18; N=903)
Gravidity (count)	3.24 (SD= 2.64; N=683)
Gender(male/female)	467/451

6.4.2 Preprocessing

As weight recordings were taken manually, preprocessing steps were carried out to ensure that neither measurement errors nor typos were considered into the analysis.

The first preprocessing step was to remove unrealistic observed weights (e.g., 10 Kg). To this end, observed weights higher than the 99th percentile reported in the WHO Growth charts[396] were removed (See Table 6.3).

Table 6.3: Maximum weights for each month based on the WHO child growth chart [396].

Month	Days after birth	Maximum weight (kg)	
		Male	Female
1st	0 to 30	5.7	5.4
2nd	31 to 60	7.0	6.5
3rd	61 to 90	7.9	7.4
4th	91 to 120	8.6	8.1
5th	121 to 150	9.2	8.7

After removing unrealistic values, the next step was to exclude outliers that could affect the regression fitting. Two different outlier problems were identified in this case. Firstly, as the majority of postnatal visits were done in the first days after birth, the local density of all observed weights after 100 days was low. These points with low density were removed using the local outlier factor (LOF), which was calculated as described in [55]. Then, using a backtracking search, the last observed point to be considered in the analysis was set as that whose LOF was lower than 1.5.

The second outlier filtering was to remove bivariate extreme points. These outliers were detected using Mahalanobis distance, in which the input matrix was composed of the observed weights and the visit days. The threshold value was set as the chi-square value with a significance level of 0.001 and a degree of freedom of two, $\chi_{0.999}^2(2) = 13.82$. Thus, any row with a Mahalanobis distance higher than the threshold was removed.

6.4.3 Fitting models

Models were fitted for different window lengths (number of days after birth) to find that one that fits the data best. The tested range was from 20 to 150 weeks, increasing by five weeks. For each window length, models were fitted using a two-stage process.

In the first stage, the Count's (Eq. 6.1), Reeds1 (Eq. 6.2), and Reeds2 (Eq. 6.3) models were applied using non-linear robust regression. For each model, nine different robust weighting functions were used to fit the model, namely: Andrews, bisquare,

Cauchy, fair, Huber, logistic, OLS (no weights), Talwar, and Welsch. Consequently, there were a total of 27 regressions, nine per each model.

The assessment of the resulted regression was based on five metrics derived from previous research showed in Table 6.1. These metrics were: i) the day in which the maximum weight loss occurred (nadir); ii) the maximum percentage of weight loss; iii) the birth weight recovery day; iv) the trend of the fitted model after recovering birth weight, and v) the rate at which weight increases after recovering birth weight.

Based on the average metrics of the Table 6.1, the reference nadir was set in 3.45 days; the percentage of weight loss in 7.817%; the recovery day in 11.15 days; the trend as monotonically increasing; and the angle at 42.26° (0.9087 Kg/month).

On the other hand, for each regression, the nadir was obtained using the first derivative. The loss percentage was calculated by comparing the origin and nadir weights. The recovery day was set as the day surpassing the origin weight. The trend after weight recovery was set as increasing if the first derivative was positive from that point to the end. Finally, the curve angle was calculated by applying the arc-tangent function to the growth rate.

The comparison between the fitted and reference vectors was performed using mean square error (MSE). Before calculating the MSE, the two vectors were normalized to avoid scale comparison problems. For each model, the weighting function with the lowest MSE was selected.

In the second stage, to reduce the effects of high individual variability reported in longitudinal studies [45, 326], the coefficients selected in the first stage were further optimized using non-linear quantile regression. The quantile regression was performed using an interior point algorithm proposed by Koenker and Park (1996)[188], setting the percentile as the median (i.e., 50th percentile), and the initial point as the coefficients found at the first stage.

6.4.4 Comparison with other models

The obtained fitted curves were compared against those obtained using parameters reported in three previous studies: Kim and Pollit (1997)[186] and Berkey (1992)[45] for Count's model, and Simondon *et al.* (1992)[335] for Count's, Reeds1 and Reeds2 models. Specifically, as these previous studies fitted a curve for each individual (longitudinal studies), the average value of their parameters was used for generating a curve from the data used in this work.

Additionally, the curves were also compared to the WHO growth charts [396], which includes newborns' data from the cities of Davis, California, USA; Muscat, Oman; Oslo, Norway; and Pelotas, Brazil; and wealthy neighborhoods of Accra, Ghana, and South Delhi, India. Thus, the fitted curve of each model was plotted against the 5th, 15th, 25th, 50th, 75th, and 95th percentiles of the WHO weight curves. This comparison allowed checking that the fitted model was growing with an inclination within the WHO percentiles.

6.4.5 Estimating birth weight

The weight gain or loss percentage function was calculated for each fitted models as:

$$g(t) = \frac{f(t) - f(0)}{f(0)}, \quad (6.4)$$

where $f(t)$ was the fitted weight value at the $t - th$ week and $f(0)$ was the fitted weight value at birth ($t = 0$). Then, the birth weight for each newborn was estimated as follows:

$$b(i) = \frac{o(i)}{1 + g(t(i))}, \quad (6.5)$$

where $o(i)$ was the observed weight for the $i - th$ newborn recorded in the postpartum visit occurred at $t(i)$ weeks, and $g(\cdot)$ is the gain/loss function (Eq. 6.4).

6.4.6 Comparison of the estimated birth weight

The estimated birth weights were compared to the Guatemalan national maternal and infant health survey [254]. However, as the data used here came from a rural community located in Tecpán, Chimaltenango, only data with similar characteristics was used for the comparison. Therefore, only 306 birth weights from rural Chimaltenango newborns, reported in the survey, were considered. From these newborns, 168 were male, whereas 138 were female.

6.4.7 Identification of low birth weight

A newborn is defined by the WHO to have LBW if the weight at birth is below 2500 g [397]. We applied this definition to the Guatemalan national maternal survey, specifically for rural Chimaltenango (the region of relevance in our study) [254]. We found that the lowest 14.3% of male newborns and 16.33% of female newborns satisfied this weight criterion.

6.5 Results

6.5.1 Preprocessed data

Figure 6.2 shows datasets before and after removing unrealistic recorded weights and outliers. In total, 17 and 28 observed weights were discarded for male and females newborns, respectively. For fitting the weight models, male newborns had a total of 450 observed weight ranging from 0 to 3.43 months (103 days), whereas female newborns had a total of 423 weights from 0 to 3.50 months (105 days). The ratio between the retained male and female weight was 1.0638, which is consistent with the 1.06 natural gender ratio reported by [138].

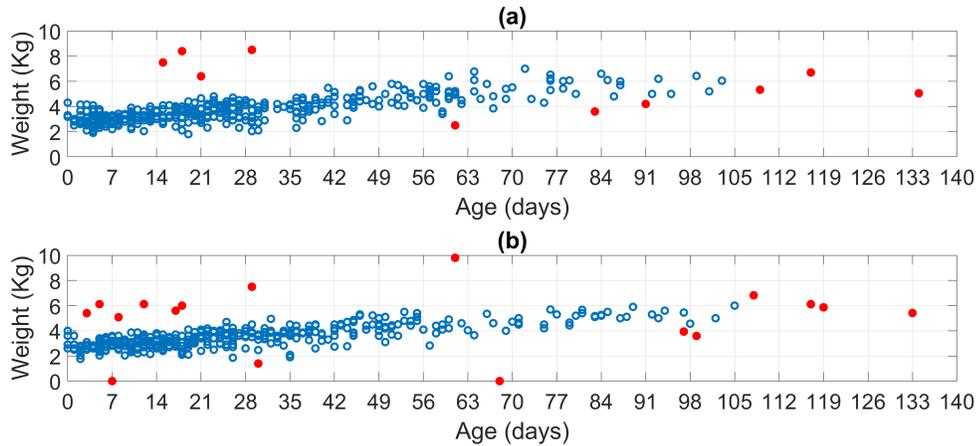


Figure 6.2: Retained (open circle) and discarded (filled red circle) observed weights for (a) male and (b) female newborns.

6.5.2 Fitted models

Post-processed data were fed into the two-stage model-fitting using different window lengths (number of days after birth). Table 6.4 shows the window length that best fitted the observed postnatal weights for each model. The percentage weight loss, the day of lowest weight (nadir), the day at which the neonate recovered their original weight and the rate of weight gain after the recovery day for each of these models were encoded in a vector and compared with the corresponding average metrics reported in previous studies (listed in Table 6.1) using the L1-norm (see final column in Table 6.4).

The Reeds2 model provided the best fit to the observations, using a window length of 60 days for males, and 45 days for females. Both genders exhibited a nadir around the fourth day after birth, with a maximum loss of 7.34% for males and 5.19% for females. Both genders recovered birth weights around the tenth day.

Table 6.5 provides the coefficients derived for the three models being compared in this work, as well as error in the regression (expressed as the MSE). For the Reeds2 model, the MSE between the fitted curve and the observed weights was 0.3 Kg^2 for males, and 0.2 Kg^2 for females.

Table 6.4: Details of the fitted curves on the observed postnatal weights, grouped by gender. The window length in days and the samples of each fitted curve are provided. The final (9th) column shows the L1-norm between a vector composed of the parameters of each model, (γ), and the target vector given by average of the metrics provided in Table 6.1, (τ), which corresponds to a nadir of 3.5 days; a weight loss at the nadir of 7.8%; a weight recovery duration of 11.2 days; and a weight gain rate of 30.3 g/day

Gender	Model	Analysis Window (days since birth)	# Samples Used	Weight Loss at Nadir (%)	Nadir (days since birth)	Recovery Day	Weight Recovery Rate (g/day)	$\ \gamma - \tau\ _1$
Male	Count's	20	209	1.40	0.12	0.25	26.4	24.50
	Reeds1	25	258	4.87	4.52	10.37	32.5	6.84
	Reeds2	60	407	7.34	4.33	11.40	35.2	5.89
Female	Count's	25	229	1.48	0.13	0.28	27.7	23.07
	Reeds1	62	374	2.23	4.00	8.90	29.1	9.47
	Reeds2	45	332	5.19	3.98	10.38	29.1	5.12

Table 6.5: Coefficients for the fitted curves shown in Table 6.4. The final (10th) column shows the MSE between the fitted curve and the observed postnatal weights.

Gender	Model	Analysis Window (day since birth)	# Samples Used	Coefficients					MSE (Kg^2)
				a	b	c	d	e	
Male	Count's	20	209	2.9690	5.6691	-6.3642			0.2213
	Reeds1	25	258	-20.5997	-6.5776	28.1120	23.6392		0.2316
	Reeds2	60	407	112.8610	17.3626	-87.6400	-152.7991	43.0524	0.3018
Female	Count's	25	229	2.8771	4.7744	-5.4227			0.2079
	Reeds1	62	374	-8.6961	-2.3848	12.9260	11.5857		0.2485
	Reeds2	45	332	116.3558	20.6091	-95.8732	-154.1696	40.7861	0.2277

6.5.3 Comparison to other models

Figure 6.3 shows the fitted Reeds2 models against previously reported coefficients [186, 45, 335] for male and female newborns. For both genders, the fitted Reeds models were more accurate than those models generated by coefficients reported in previous research.

Likewise, Figure 6.4 compares the fitted models to those corresponding to the weight percentiles of the WHO weight chart. For both genders, the Reeds models grew along with the 25th percentile, which is expected as WHO weight chart includes populations of industrialized countries, as well as affluent neighborhoods in India and Ghana.

6.5.4 Estimation of birth weight and identification of low birth weight

The Reeds2 models were used to estimated birth weight following steps presented in subsection 6.4.5. Part (a) of the Figures 6.5 and 6.6 shows the distributions of

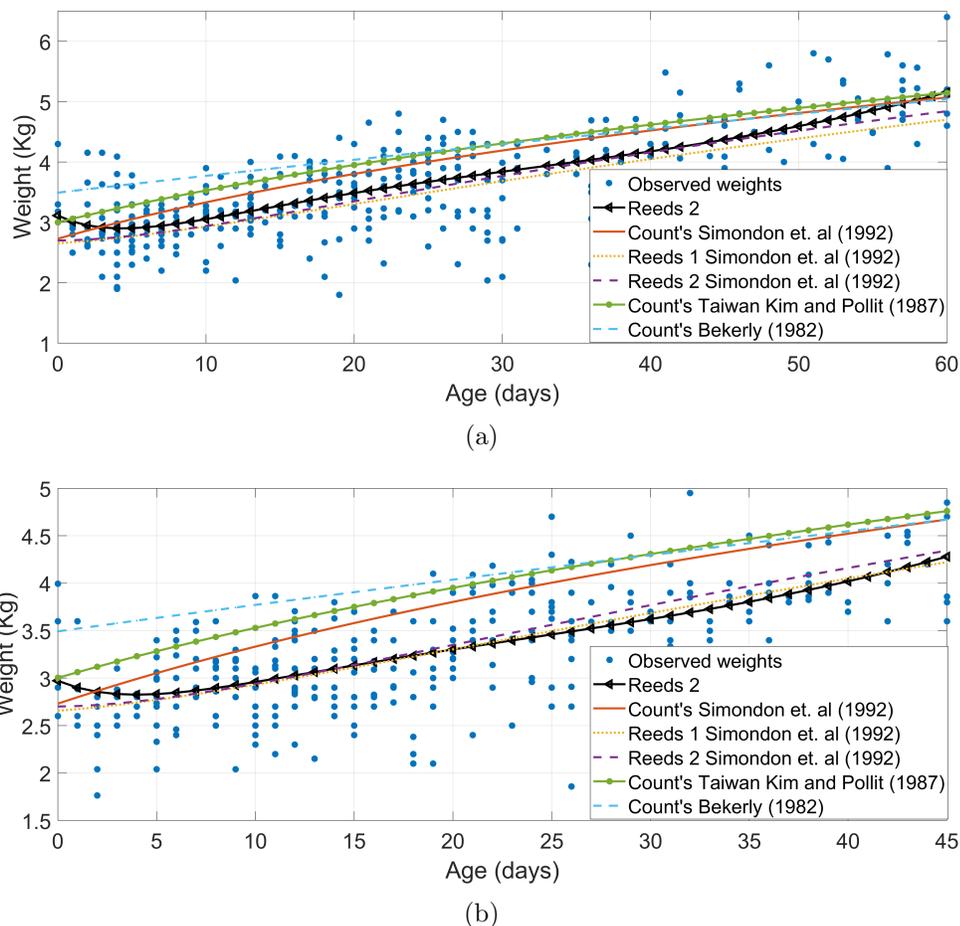


Figure 6.3: Reeds2 models fitted for (a) male and (b) female newborns against models generated using coefficients reported by Kim and Pollit (1987)[186] in a Taiwanese population, Berkey (1992)[45] in a Boston hospital sample, and Simondon *et al.* (1992)[335] in a Congo rural community.

estimated birth weight for male and female newborns, respectively. For both genders, the majority of the estimated birth weights were close to 3 Kg, being lower for females. Moreover, the corresponding LBW percentile threshold was located in the estimated birth weights. For male newborns, the weight threshold was found at 2.64 Kg, whereas for females, it was 2.57 Kg.

Additionally, Figures 6.5 and 6.6, in part (b), shows the distributions of estimated birth weights and those reported in the 2015 Guatemalan national survey for rural Chimaternalgo newborns [254]. Estimated weight distributions were similar to the surveyed distributions. In fact, in Table 6.6, it is shown that for any gender, there

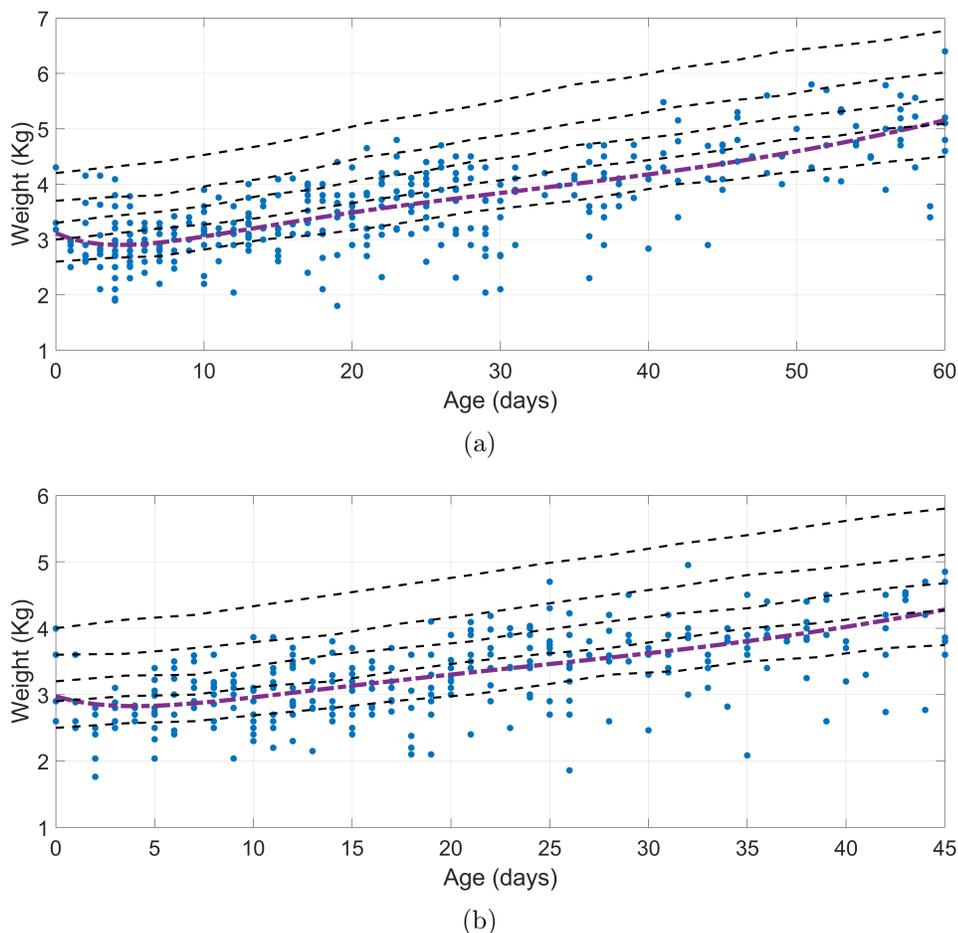


Figure 6.4: Reeds2 models fitted for (a) male and (b) female in observed weights after birth against WHO weight growth chart percentiles. The percentiles showed for the WHO weight dataset corresponds to the 5th, 25th, 50th, 75th, and 95th.

was no statistically significant difference between the medians of the estimated birth weights and the reported weights for rural Chimatenalgo newborns.

Table 6.6: Descriptive statistics of the estimated birth weights and the 2015 Guatemalan national survey reported newborns of rural Chimatenalgo [254]. Last two columns show the p -value of: (i) a two-sample t -test; and (ii) a paired two-sided Wilcoxon signed rank test. Null hypothesis tests were whether the two weight distributions have equal mean and median, respectively.

Gender	Distribution	mean	standard	kurtosis	skewness	t-test	Wilcoxon signed-rank
		(Kg)	deviation (Kg)			p-value	p-value
Male	Estimated birth weights	3.10	0.47	3.60	-0.32	0.3765	0.7479
	2015 Guatemalan national survey	3.15	0.64	3.89	0.28		
Female	Estimated birth weights	2.95	0.43	3.31	-0.32	0.1913	0.1495
	2015 Guatemalan national survey	3.02	0.60	4.43	-0.19		

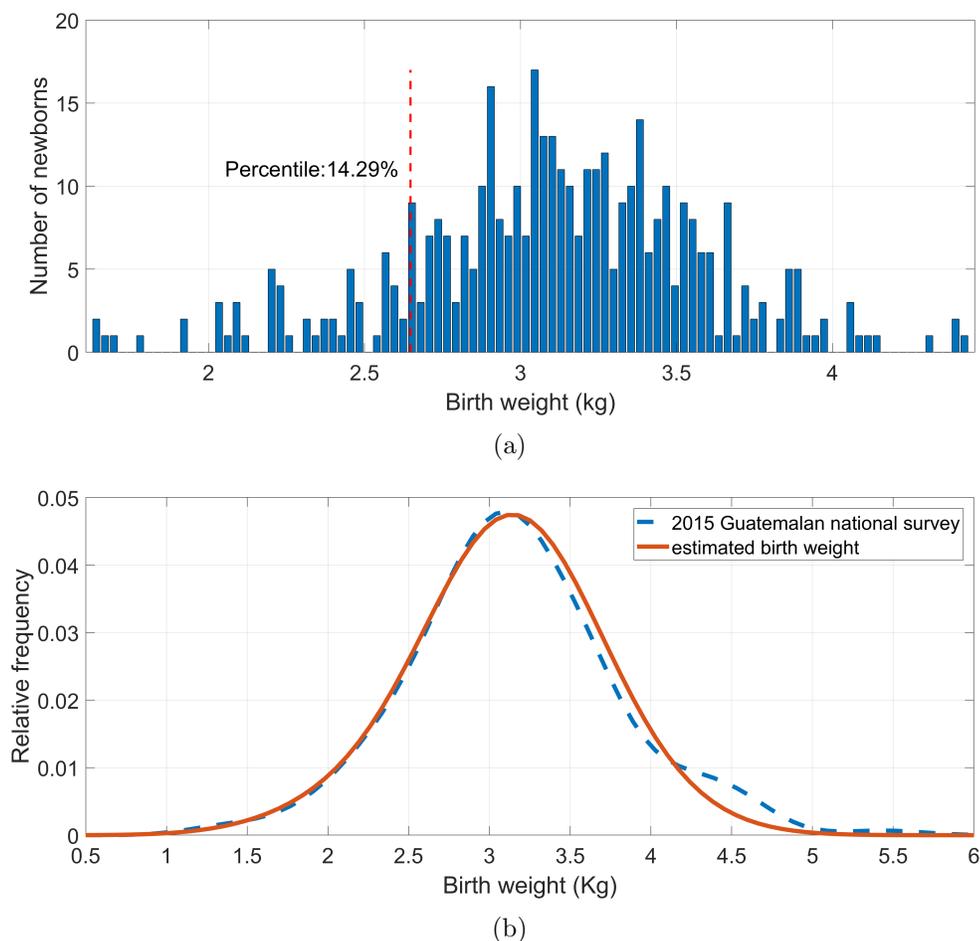


Figure 6.5: Male newborns' birth weights estimated using the fitted second-order Reeds model. (a) Histogram of estimated birth weights with the threshold for identifying low birth weight (vertical dotted line), corresponding to the 14.29 percentile, or 2.64 Kg. (b) Comparison of the distribution of estimated birth weights (red) with the measured birth weights in the 2015 Guatemalan national survey for male newborns of rural Chimatenalgo (blue dashed line).

6.6 Discussion

The fitted weight curves presented here indicates that it is possible to estimate birth weight from observed weight recorded days or months after birth. Notably, our two-stage fitting steps were able to describe a typical newborn weight pattern in a dataset composed of only one sample per newborn rather than multiple points per subject as described in previous research works [248, 116, 393, 335, 45, 186, 97, 282, 410, 80, 19, 276]. This suitable fitted pattern it is likely explained by the homogeneity of subjects used in this work, in which maternal nutrition, and delivery and feeding methods are

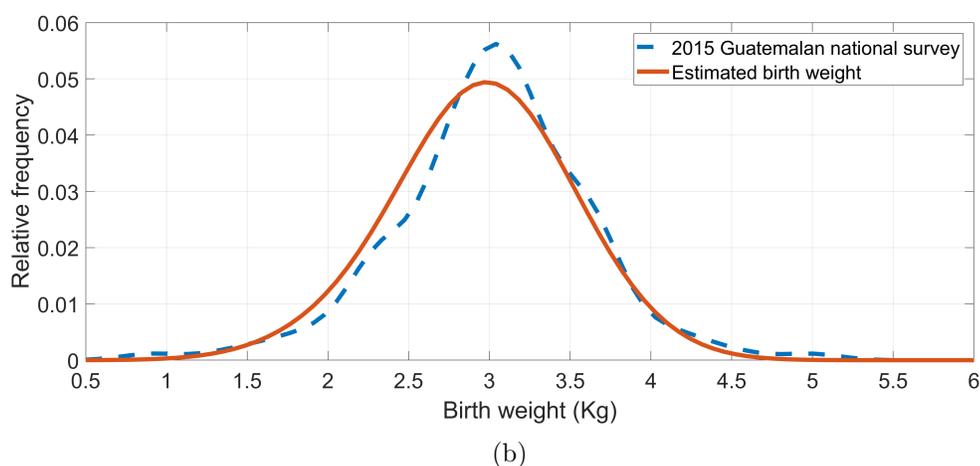
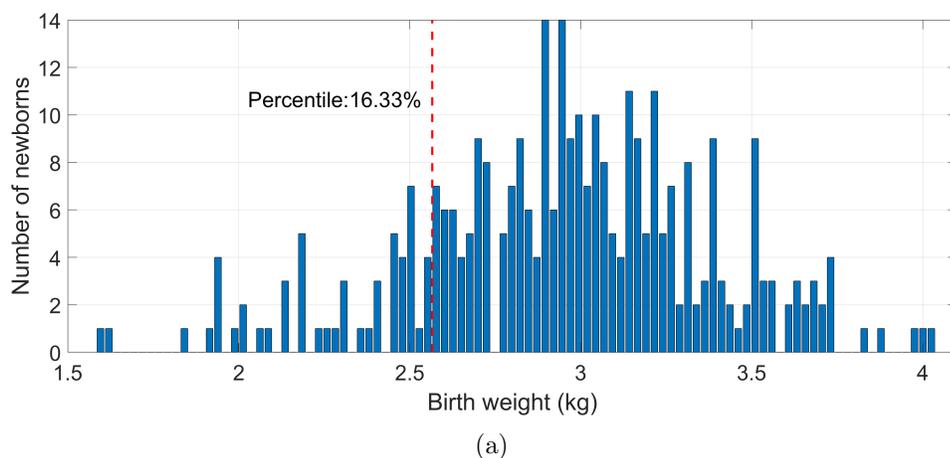


Figure 6.6: Female newborns' birth weights estimated using the fitted second-order Reeds model. (a) Histogram of estimated birth weights with the threshold for identifying low birth weight (vertical dotted line), corresponding to the 16.29 percentile, or 2.57 Kg. (b) Comparison of the distribution of estimated birth weights (red) with the measured birth weights in the 2015 Guatemalan national survey for female newborns of rural Chimaternalgo (blue dashed line).

common among them.

Within the infant weight models, the Reeds2 model fitted the observed data best, thereby supporting the previous finding reported by Simondon *et al.* (1992)[335], in which Reeds models resulted in the lowest residuals. As the Reeds2 model has two inflection points [46], the curve was able to precisely adjust to the different infant weight patterns than the other models.

In comparison with previous longitudinal (see Table 6.1), the Reeds2 model obtained similar metrics for both genders, showing a maximum weight loss within the third and fifth day, and recovery the birth weight by ten days after birth. Although

the weight percentage loss was lower than the expected 7% for females, it can be explained by the fact that the model was fitted for all subjects, thereby introducing inter-variability in the fitted curve [282, 326]. Nevertheless, the achieved percentage loss, as well as the other model criteria, were within the full ranges reported in previous studies.

The median quantile regression applied to the Reeds model allowed estimating a more robust and less biased trend of the population. Specifically, the second-order Reeds model grew with a feasible rate along with the 25th percentile of the WHO growth chart (Figure 6.4). Although the fitted median quantile regression was lower than the 50th CDC percentile, there is not any inconsistency as it is expected that Guatemalan newborns have a lower weight than those of a well-equipped-resource environment, such as the US, Norway, and high-income population in New Delhi [264]. In fact, the WHO 50th quantile birth weight is close to 3.5 Kg, which is the average for European-inheritance newborns.

The average of the estimated birth weights, for both male and female, was consistent with the figures reported by the 2015 Guatemalan national survey in rural Chimaternalgo [254], varying less than 20 g. Likewise, the estimated averages were also close to the 2983.2 g reported by Bose *et al.* (2015)[52] for a Guatemalan sample of 30262 deliveries. Therefore, for both gender, our approach estimated birth weights in a feasible range (see Table 6.6).

Interestingly, by mapping rural Chimaternalgo LBW percentiles, we obtained a weight threshold for identifying LBW in our estimations. This identification is relevant for TBAs as they may be able to detect any potential LBW cases. Thus, newborns could receive early treatment, thereby reducing short-and long-term complications.

The main limitation of this work is that the observed dataset was not enough to carry on a longitudinal study. Therefore, the weight pattern obtained is an approxi-

mation and does not represent a complete characterization of the weight gain pattern of the rural Guatemala population used here. Nevertheless, the simple method used here provides a straightforward way to estimate birth weight, thus helping to identify low birth weight newborns on late postpartum visits.

Future research should perform a longitudinal study by periodically weighing newborns from birth to one year old. Thus, the two-stage method introduced here could be fitted for each newborn, resulting in a more accurate characterization of the newborns' weight curve for the Guatemalan highland community.

6.7 Conclusion

In LMICs, timely recording of birth weight is often difficult. This work introduces a simple approach for estimating birth weight from observations recorded after birth. The estimations were consistent with infant weight patterns reported in previous studies conducted in Guatemalan communities. By weighing infants several days after birth, it may, therefore, be possible to identify LBW newborns and initiate treatment to reduce short-and long-term complications. Nevertheless, we note that the neonatal assessment needs to occur in the first few days of life to accurately estimate a birth weight relevant for tracking neonatal health [205].

Chapter 7

Estimating gestational age from 1D-DUS and maternal blood pressure recordings

7.1 Abstract

In-utero progress of fetal development is normally assessed through manual measurements taken from ultrasound images, requiring relatively expensive equipment and well-trained personnel. Such monitoring is therefore unavailable in low- and middle-income countries (LMICs), where most of the perinatal mortality and morbidity exists. The work presented here attempts to identify a proxy for IUGR, which is a significant contributor to perinatal death in LMICs, by determining gestational age (GA) from data derived from simple-to-use, low-cost one-dimensional Doppler ultrasound (1D-DUS) and blood pressure devices. A total of 114 paired 1D-DUS recordings and maternal blood pressure recordings were selected, based on previously described signal quality measures. The average length of 1D-DUS recording was $10.43 \text{ mins} \pm 1.41 \text{ mins}$. The min/median/max systolic and diastolic maternal blood pres-

tures were 79/102/121 and 50.5/63.5/78.5 mmHg, respectively. GA was estimated using features derived from the 1D-DUS and maternal blood pressure using a support vector regression (SVR) approach and GA based on the last menstrual period as a reference target. A total of 50 trials of five-fold cross-validation were performed for feature selection. The final SVR model was retrained on the training data and then tested on a held-out set comprising 28 normal weight and 25 low birth weight (LBW) newborns. The mean absolute GA error with respect to the last menstrual period was found to be 0.72 and 1.01 months for the normal and LBW newborns, respectively. The mean error in the GA estimate was shown to be negatively correlated with the birth weight.

Thus, if the estimated GA is lower than the (remembered) GA calculated from last menstruation, then this could be interpreted as a potential sign of IUGR associated with LBW, and referral and intervention may be necessary. The assessment system may, therefore, have an immediate impact if coupled with suitable intervention, such as nutritional supplementation. However, a prospective clinical trial is required to show the efficacy of such a metric in the detection of IUGR and the impact of the intervention.

7.2 Introduction

Estimation of fetal gestational age (GA) provides important information throughout pregnancy, such as delivery scheduling, growth disorder detection, and preterm newborns management [7]. Thus, GA estimation can assist in detecting issues leading to perinatal mortality and morbidity [304, 181]. This detection is particularly needed in low-and middle-income countries (LMICs), which account for approximately 98% of all reported perinatal deaths worldwide, largely due to gestational developmental issues [412].

In high-income countries, clinical teams generally use ultrasound images to estimate GA, as well as any structural abnormalities [226]. These GA estimations are based on a variety of fetal measurements, such as biparietal diameter (BD), crown-rump length (CRL), head circumference, abdominal circumference, and femur length [226]. However, in LMICs, the access to ultrasound imaging is limited, and almost unavailable in rural areas, due to the high cost of the medical equipment, the expenses for maintenance, and the requirement of skilled medical staff [399]. Hence, low-cost alternative methods for dating gestation are used in LMICs.

A common low-cost method used for GA estimation is the last menstrual period (LMP), in which a 28-day menstrual cycle is assumed. Although previous studies have criticized LMP due to the inconsistency in the menstrual cycle length [95], and the difficulty to recall the day of the last menstrual period [20], the LMP method has shown to be a somewhat useful method for LMICs, particularly in rural areas lacking medical equipment. In fact, [265] compared 171 GA estimations based on LMP collected in rural Guatemala with GA estimations given by BD, reporting that GA estimations by the LMP were within ± 14 days of the BD estimations for 94% of the cases.

GA estimation can assist in the assessment of intrauterine growth restriction (IUGR), which has a prevalence varying between 9 to 11% in LMICs [89, 205]. Specifically, IUGR is assessed by comparing the estimate of GA with the symphysis-fundal height (SFH) measurement [403]. For fetuses growing normally, from 24 weeks of gestation, the SFH measurement (L_{sfh}) in centimetres should correspond to the number of weeks of gestation ± 2 cm. When $L_{sfh} < N - 2$, where N is the number of weeks since the last menstrual period, the fetus is suspected to be IUGR [287]. However, the SFH method lacks significant evidence to recommend its widely use in LMICs [403]. Moreover, previous studies have noted that the SHF has exhibited a large error of ± 6 weeks for estimating GA [139]. New approaches are, therefore, still needed to

provide repeatable and low-cost assessment for detecting abnormal growth in settings in which ultrasound images, taken by trained operators, are not available.

In this work, we propose an alternative approach for GA estimation to provide a proxy for assessing fetal development and identifying possible cases of IUGR for a Guatemalan rural population, in which ultrasound imaging is not affordable and the SFH is not accurate. Our approach estimates GA using fetal heart rate variability (fHRV) indexes and maternal hemodynamics derived from one-dimensional Doppler ultrasound (1D-DUS) and maternal blood pressure, respectively. Data were acquired during routine perinatal check-up visits by traditional birth attendants (TBAs) using a low-cost Doppler transducer and a self-inflating blood pressure device [354, 351, 232, 231]. These features were used to build a machine learning algorithm to estimate GA. We hypothesized that if the estimated GA is lower than the GA calculated from last menstruation, then this could be interpreted as a potential sign of IUGR based on low birth weight (LBW), and referral and intervention may be necessary.

7.3 Background

Fetal heart rate is regulated by the Autonomic Nervous System (ANS) [323, 388], which in turn modifies FHR dynamics over the course of pregnancy. In particular, FHR variability evolves over the course of pregnancy and may reflect the maturity of the ANS, and thus may indicate the fetal GA. [386] reported that fHRV, as observed from traces taken from 61 pregnant women without complications, increases during gestation. In particular, they noted that short term variability increased during the last trimester, whereas long term variability exhibited the largest increases in the early gestational period. Figure 7.1 shows an example of how FHR changes across gestation, as reported in [386].

Based on FHR, previous studies have shown a correlation between GA and mark-

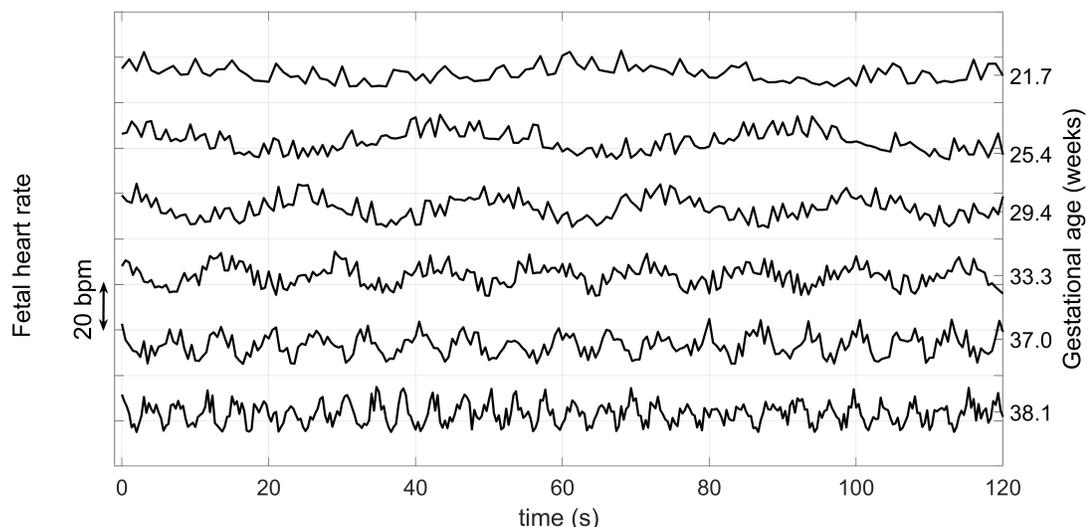


Figure 7.1: Variation of the FHR from 22-38 weeks during pregnancy. Note that the vertical axis has an arbitrary offset. Average FHR does not drop by such a large amount each week during pregnancy, but rather it drops on average by about 15 BPM from week 25 to week 40 [180]. Adapted from [386]

ers derived from fHRV. Linear metrics such as the mean of R-R interval (mRR), the standard deviation (sdRR) and root mean square of successive differences (rmsdRR) positively correlated with GA for both genders [198]. Non-linear metrics, such as approximate entropy (ApEn), Lyapunov exponent, tone-entropy and generalized mutual information, have also been linked to fetal maturation [380, 156, 184]. Additionally, [379] and [332] reported that power in the 0.003-1.0 Hz frequency band vary during pregnancy.

Initial works on GA estimation aimed to find a relation between GA and FHR-based metrics using univariate regression [156]. More recently, some works have aimed to improve the characterizing of FHR by incorporating multivariate and more complex methods. In particular, [360] extracted features from 359 high resolution fetal magnetocardiographic (fMCG) recordings, lasting at least 20 minutes. The researchers implemented an algorithm to extract non-active portions of the recording and calculated both linear and non-linear metrics of fHRV from the these *quiet* periods. Results showed that entropy and skewness were more highly correlated with GA than those

obtained by traditional linear HRV metrics. However, this approach requires high temporal and spatial resolution data acquired from costly and non-portable equipment, making its use in LMICs impractical.

In earlier work, [235] estimated GA for 57 fetuses using a step-wise regression based on cardiac wall intervals derived from one-dimension Doppler ultrasound signal (1D-DUS) and fECG signals recorded in a Japanese hospital. The estimated GAs were compared to the GA derived from CRL, achieving a mean square error of 3.8 and 5.1 weeks for cardiac intervals and fHRV parameters, respectively. In further work, [237] improved the estimation accuracy by incorporating 1D-DUS and fECG quality assessment algorithms to filter poor quality signals. As a result, the step-wise regression achieved a mean absolute error (MAE) of 4.7 weeks from fHRV parameters, and 2.7 weeks when including the cardiac intervals metrics. Although this latter method achieved comparable results to Doppler imaging based estimations, it required two sources, 1D-DUS and fECG signals, which increases costs and complicates implementation, particularly in LMICs [354].

In addition to FHR indexes, maternal blood pressure is also a relevant metric for GA estimation. Previous works have reported that maternal systolic and diastolic blood pressure increases throughout pregnancy [348, 315, 178, 297]. However, despite the correlation between GA and maternal blood pressure, no research has included maternal blood pressure in regression models to estimate GA. We note that extreme blood pressures may be indicative of pre-eclampsia, or other gestational issues. It is therefore important to treat these separately.

7.4 Methods

7.4.1 Databases

Collection of the data

Data used in this work were collected as a part of a randomized control trial conducted in rural highland Guatemala in the vicinity of Tecpan, Chimaltenango. This program was approved by the Institutional Review Boards of Emory University, the Wuqu' Kawoq | Maya Health Alliance, and Agnes Scott College (Ref: IRB00076231 - 'Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala') and registered on ClinicalTrials.gov (identifier NCT02348840). In the trial, traditional birth attendants were trained to use a mobile mHealth system to record perinatal information during approximately monthly visits during the second and third trimesters. More details on the design and implementation of the mobile mHealth system, and the training of the TBAs can be found in [354, 351] and [232, 231].

The perinatal care program included both prenatal and postpartum visits. In the prenatal visits, TBAs recorded GA in months by counting the number of whole months since the last menstrual period. The GA was recorded in months instead of weeks to reduce measurement errors since usually patients attended in this project forgot the specific date of their last menstrual period, and very few received an early obstetrical ultrasound for more accurate dating [232]. During the visit, the TBA also recorded 1D-DUS signals and maternal blood pressure using the mobile mHealth system [354, 351, 232, 231]. The 1D-DUS signals were recorded using a Doppler transducer (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 maternal blood pressure was taken in a supine position from both arms using a self-inflating blood pressure device calibrated for pregnancy.

In the postpartum visits, TBAs recorded the newborn's birth date, sex, current weight, length, and head size. These postpartum visits could occur days or months after birth since sometimes it was difficult to follow up on the patients. The birth weight was then estimated using a Reed2 second-order model [46] fitted on 917 observed postnatal weights using an approach we have previously described in [372]. A weight threshold was used to classify the estimated birth weights as low or normal. This threshold was defined by first finding the percentile corresponding to ≤ 2.5 Kg in a Guatemalan national maternal survey for the region of relevance in our study [254]. We found that the lowest 14.3% of male newborns and 16.33% of female newborns satisfied this weight criterion. These percentiles were then located in our estimated birth weight distribution to determine the corresponding LBW threshold. This corresponded to 2.64 Kg for males and 2.57 Kg for females.

Assumption of the study

In this work, a newborn was considered as a possible case of IUGR if their estimated birth weight was below the threshold discussed above. This assumption is based on the fact that LBW could be a consequence of either preterm birth (< 37 weeks) or small-for-gestational-age (SGA). However, in LMICs, around 60% of LBW newborns are SGA [205], and the main reason for SGA in LMICs is IUGR [89, 205].

Data Inclusion Criteria

Prenatal visits were included if they contained both blood pressure pictures and 1D-DUS recordings with some conditions. For the blood pressure, the numbers had to be readable on the photograph of the blood pressure device. Also, the difference between the right and left arm measurements had to be lower than 15 mmHg, thus discarding any spurious measurements. Finally, possible preeclampsia patients were discarded when systolic or diastolic blood pressure was higher than a threshold. This threshold

was defined at 130 and 80 mmHg for SBP and DBP, as is suggested for measurements taken in spine position [263, 187, 71].

The conditions for including the 1D-DUS recording were based on length and quality. The minimum length was fixed at ten minutes since earlier work suggests that this is the required length to extract fHRV indexes, such as baseline, accelerations, and decelerations [99].

In addition to the length, the quality of a 1D-DUS recording was also considered as an inclusion criterion. The 1D-DUS quality was assessed using a window of 3.75 *s* and a sliding window of 250 *ms*. For each 3.75-second window, 16 features were extracted, including Wavelet percentage energy in the range 250 to 2000 Hz, Mel-frequency cepstral coefficients (MFCC), and power spectrum ratios on electrical interference frequency ranges. The features were fed into a classifier composed of a logistic regression and a multiclass support vector machine to classify the 3.75-second window into good quality, interference, silence, talking in the background, or low signal to noise ratio. More details of the quality assessment method can be found in [373, 374].

Based on the length and quality criteria, a 1D-DUS recording was only included in this present work if it lasted more than 10 minutes, and at least 50% of its 3.75-second windows were labeled as good quality.

Final data set

After applying the inclusion criteria, the final dataset comprised 167 visits from 153 non-preeclampsia women who were pregnant with singleton fetuses. From these patients, 142 gave birth to normal weight singletons, whereas 24 gave birth to LBW newborns, based on our thresholds defined for the study population [372] (see subsection 7.4.1).

Table 7.1 shows demographics of the patients. The male/female ratio was higher

in the LBW group than in the normal birth weight group. On the other hand, the maternal age and the number of previous pregnancies (gravidity) were higher in the normal weight group.

Table 7.1: Average demographics for the data used in this study. For each metric, the standard deviation and the number of patients available for that variable are shown in parenthesis.

Variable	Normal birth weight	Low birth weight
Patients (count)	129	24
Newborn gender (male/female)	56/73	17/7
Birth weight (Kg)	3.1 (SD=0.3;N=129)	2.3 (SD=0.4;N=24)
Maternal age (years)	27.0 (SD=6.3;N=123)	24.5 (SD=6.7;N=22)
Gravidity (count)	3.4 (SD=2.5;N=96)	2.3 (SD=3.1;N=16)

Table 7.2 shows the distribution of the GA based on the last menstrual period (LMP) method. Visits ranged from the sixth to the ninth month of pregnancy, focusing mainly on the third trimester.

Table 7.2: Number of visits per gestational age (GA) taken with the last menstrual period (LMP) method.

Gestational age (months)	Normal weight	Low birth weight
6	9	3
7	38	7
8	37	5
9	58	10
total	142	25

7.4.2 Deriving the FHR signal

Extracting fetal heart rate from 1D-DUS

Each 1D-DUS recording was analyzed using a window of 3.75 s and a sliding window of 0.25 s. The window length was set at 3.75 s since it is the usual length for computerized analysis of fetal non-stress tests based on the Dawes/Redman criteria [85, 279]. The selection of the sliding window was based on the desired sampling

frequency, namely 4 Hz. This sampling frequency has been shown to be sufficient for digital cardiotocography [309], and corresponded to a Nyquist frequency of 2 Hz, thus allowing the extraction of spectral metrics in the range 0.03-1 Hz.

For each 3.75-second window, the fetal heart rate (FHR) was estimated auto-correlation (AC)-based method, which we previously introduced in [375]. Specifically, the method detects the fundamental period of the envelope of the 3.75-second window by applying auto-correlation, and then the FHR is estimated by dividing 60 between the fundamental period in seconds. More details of the FHR estimator are found in [375].

In addition to estimate the FHR, the quality of the 3.75-second window was also assessed and stored for further preprocessing steps. The quality was assessed using the method presented in [374] (see subsection 7.4.1).

Preprocessing of estimated FHR signal

Since 1D-DUS recordings are prone to noise, any given 3.75-second window of Doppler data may lead to an unreliable estimate. Two steps then assessed the reliability of the extracted FHR. Firstly, as recommended in [271], we removed FHR estimates that were not within the 65-175% range of the average of the previous two estimates. Secondly, we removed 3.75-second windows classified as something else other than good quality.

Each discarded value was replaced by the linear interpolation between its previous point and the next stable segment. A stable segment was defined as a region of five adjacent points for which the FHR estimate did not vary by more than ten beat per minute (BPM).

Calculation of the baseline, acceleration and deceleration

The baseline was determined using an algorithm proposed by [22], which is an improvement of those proposed by [83] and [228]. Specifically, a filter bank was applied to the 4-Hz FHR time series to attenuate any accelerations or deceleration.

Following the work of [22], accelerations and decelerations were detected for each valid one-minute segment of the baseline. A valid segment was determined by computing a FHR histogram using a bin width of 10 BPM. If the most frequent bin of the histogram contained more than 40% of the values, the one-minute baseline segment was considered valid.

When a one-minute baseline segment was determined to be valid, acceleration and deceleration intervals were identified following Dawes criteria [83]. Namely, an acceleration was defined to be a section of data for which the FHR was higher than the baseline for at least 15 *s* and at least one sample was 15 BPM or more above the baseline. Similarly, a deceleration was defined where the FHR remained below the baseline FHR for at least 15 seconds, and at least one sample was 15 BPM below the baseline. More details of the algorithm can be found in [22].

7.4.3 Features used for gestational age estimation

Based on previous works presented in section 7.3, a total of 37 features relevant for estimating GA were extracted from the 1D-DUS and blood pressure device captured at the perinatal visits.

The features derived from 1D-DUS recording were calculated using the FHR time serie (see section 7.4.2). Since the RR-interval sequence is necessary to estimate fHRV metrics, the FHR series was converted into a interbeat sequence as:

$$T(i) = 60000/S(i), \quad (7.1)$$

where $S(i)$ is the FHR at the $i - th$ second. Using the FHR time serie and the interbeat sequence, three different type of features were calculated, namely: linear, non-linear and complexity, and spectral. From the maternal blood pressure and heart rate, seven hemodynamics formulas were calculated.

Table 7.3 shows the total features extracted, discriminated by source and type. Next subsections presents the calculation of all the features.

Table 7.3: Features used for GA estimation. The features included temporal features as basal heart rate; STV; LTV; II; LTV; STV/LTV ratio; ACC; DCA; the number of accelerations per minute; and the variance, mean, standard deviation, skewness, kurtosis, PNN5, root mean square difference of successive interbeat sequence. Features also included complexity features and spectral features as ApEn, fractal dimension, Lyapunov exponent, LF, MF, HF, LF/(MF+HF) ratio, Tone-Entropy, and Generalized mutual entropy. Finally, maternal hemodynamic formulas were also included.

Source	Type	Name
Doppler transducer	Linear	STV, LTV, II, LTI, basal FHR, STV/LTV, ACC, DAC, #accelarations/minute
	Non-linear& Complexity	varIS, mIS, stdIS, rmssdIs, skewnessIs, kurtosisIs, stdIS/rmssdIs, PNN5
	Spectral	ApEn, Fractal Dimension, Lyapuno exponent Entropy, Tone, GMI
Self-inflating blood pressure device	Raw	SBP, DBP, MHR
	Hemodynamics formulas	PP, MAP, CO, RPP,SI, MSI, SV

Linear features from the interbeat sequence

The baseline, acceleration, and deceleration were used to calculate variability metrics previously reported to monitor fetus wellbeing during pregnancy [216, 109, 331].

Since commercial cardiotocography (CTG) devices calculate linear fHRV metrics by buffering and averaging ten consecutive FHR readings [332], we also reduced the sampling frequency to 0.4 Hz for linear fHRV metrics. To that end, for each minute, the 240 FHR values were reduced to 24 by averaging ten consecutive values without overlapping. The corresponding 24 FHR values per minute were converted to 24 interbeat sequence values using Eq. 7.1.

Short term variability Short term variability (STV) quantifies variability within a one-minute length window. Using the 24 interbeat sequence values of each minute, the one-minute STV was calculated as:

$$STV_j = \frac{\sum_{i=2}^{24} |T_j(i) - T_j(i-1)|}{23}, \quad (7.2)$$

where $T_j(i)$ is the i -th sample of the j -th minute of the interbeat sequence.

The STV of the total interbeat sequence was calculated by averaging the one-minute STV, excluding those minutes in which decelerations occurred [83].

Interval Index The interval index (II) is an alternative index measuring long term variation that takes into account the standard deviation of the interbeat sequence. It was calculated for each minute as:

$$II_j = \frac{std[|T_j(2) - T_j(1)|, |T_j(3) - T_j(2)|, \dots, |T_j(24) - T_j(23)|]}{STV_j}, \quad (7.3)$$

where std is the standard deviation of the absolute difference of successive values of the j -th minute of the interbeat sequence. The one-minute Interval Index (II) values were averaged to obtain the total value of the interbeat sequence.

Long term variability Long Term Variability (LTV) is the range between the highest and the lowest interbeat sequence for each minute [99]. It was calculated for each minute as:

$$LTV_j = max[T_j] - min[T_j], \quad (7.4)$$

where T_j is the j -th minute of the interbeat sequence. The one-minute LTV values were averaged to obtain the total value of the interbeat sequence.

Long Term Irregularity Long Term Irregularity (LTI) measures variability over longer time scales. LTI was calculated for a window length of 3 minutes, as was recommended by [331]. LTI is defined as the interquartile range of the following distribution:

$$m_k = \sqrt{\sum_{i=2}^{72} T_k^2(i) + T_k^2(i-1)}, \quad (7.5)$$

where $T_k(i)$ is the i -th value of the k -th three-minute segment of the interbeat sequence. Note that the summation index ranged from 2 to 72, as the length for calculation was three minutes. The total LTI of the recording was taken as the mean of the three-minute values.

Basal fetal heart rate The basal heart rate was estimated on the original FRH sequence following the procedure explained in [22]. First, FHR segments in which accelerations and decelerations occurred were discarded, and then, the kept FHR values were used to compute a histogram. The center of the most frequent bin was selected as the basal fetal heart rate.

Number of accelerations per minute As the acceleration of the heart rate is associated with the maturity and fetal wellbeing [15], we also included the number of accelerations per minute as a feature. This feature was calculated by dividing the total number of accelerations of the interbeat sequence over the length in minutes of the sequence.

Acceleration average capacity and deceleration average capacity Acceleration average capacity (AAC) and deceleration average capacity (DAC) were calculated from the phase-rectified signal (PRSA) of the interbeat sequence by following steps presented in [163]. To that end, two parameters were defined. The first parameter is the filter condition, M , used to find anchor points. The second parameter is a time

window (length $2L$), used to define a window around the anchor points.

Here, the parameters, M and L , were both set to one, which we found to be the optimal values for FHR derived from 1D-DUS signals in earlier work [354, 355].

For AAC, an anchor points was found when the average of M points is lower than 5% of the average of the M successive points. For DAC, an anchor point corresponded to a decrease of more than 5% between successive averaged values. Thus, the AAC and the DAC were obtained by aligning their corresponding anchor points and averaging the windows. More details are found in [163].

Statistical movements of the interbeat sequence As reported by [198], the mean (mIS), standard deviation (stdIS), variance (varIS), skewness (skewnessIS), kurtosis (kurtosisIS), and root mean square of successive differences (rmssdIs) of the interbeat sequence were calculated. We also calculated the ratio between the standard deviation and root mean square of successive differences (stdIS/rmssdIs).

Additionally, the fraction of consecutive beats that differ by more than 5 milliseconds (PNN5) was also calculated. PNN5 was calculated using an open source cardiovascular toolbox introduced in [383].

Non-linear and complexity features

Non-linear and complexity metrics were also extracted from the FHR time series as suggested in [380, 332, 156, 331, 184]. Six different features were calculated:

- Approximate entropy (ApEn). ApEn was calculated with the cardiovascular toolbox [383], setting the m and r parameters in 2 and 0.1 of the standard deviation of the input signal.
- Fractal dimension. The fractal dimension was calculated using the Higuchi's algorithm [148], setting the interval parameters as 5.

- Lyapunov exponent. The Lyapunov exponent, previously reported to be negative correlated with GA [380], was calculated following steps presented in [310] setting embedding dimension and lag parameters as 1 and 2, respectively.
- Entropy and Tone. Tone-entropy has shown to assess the development of ANS throughout pregnancy [184]. Tone is the average of the percentile difference of successive beat intervals (PI). Entropy is calculated by using the Shannon formula [328] on the PI distribution.
- Generalized mutual information (GMI). The general mutual information was calculated following steps in [156], setting dimension parameter at 3 and delay parameter at 1.

Frequency features

Frequency were also extracted from the FHR time series as suggested in [379, 332, 331]. The power spectrum density of the FHR time series was averaged in three frequency bands: the low, middle, and high bands. The ranges for these bands were defined as 0.03-0.15 Hz for low frequency (LF), 0.15-0.5 Hz for medium frequency (MF), and 0.5-1 Hz for high frequency (HF). These power spectral ratios were extracted using the open source cardiovascular toolbox [383].

Maternal blood pressure and hemodynamic formulae features

The maternal systolic blood pressure (SBP) and diastolic blood pressure (DBP), and the maternal heart rate (MHR) measurements from the blood pressure device were used as features, as well as hemodynamic formulas derived from them. Since the SBP, DBP, and MHR were taken for both patient's arms, these values were averaged values. These were then used to calculate hemodynamic formulae (see Table 7.4), which have been reported to vary throughout pregnancy [348, 315, 297, 178].

Table 7.4: Detail of maternal hemodynamic formulae calculated using the SBP, DBP and MHR taken with the self-inflating blood pressure device.

Metric name	Formula	Reference
Pulse pressure (PP)	$SBP - DBP$	[350]
Mean arterial pressure (MAP)	$(SBP + DBP \times 2)/3$	[350]
Cardiac output (CO)	$MHR \times PP \times 0.002$	[149]
Rate pressure product (RPP)	$MHR \times SBP$	[307]
Shock index (SI)	MHR/SBP	[337]
Modified Shock Index (MSI)	MHR/MAP	[337]
Stroke volume (SV)	CO/MHR	[350]

7.4.4 Estimation of gestational age

All of the features were extracted for both the 129 normal birth weight and 24 LBW newborns at each stage of pregnancy for which data was available. The GA estimation model was training only with visits of newborns with normal birth weights because previous research has reported that LBW fetuses have discrepancies in their GA estimations from fHRV [237]. However, the features derived from the recordings of the LBW newborns were used later to test the model's ability to estimate GA.

The 129 normal weight patients were split into training and test sets. The number of patients for the test set was selected to be proportional to the LBW newborn set. A Wilcoxon rank-sum hypothesis test (two-sided; $\alpha = 0.05$) was applied in order to test whether there were statistically significant differences between the training and test sets for all the values of the 37 features. (if a statistically significant difference was found, the subjects were randomized again.)

The training set comprised 104 newborns, from which 95 had one visit, eight had two visits, and one had three visits, giving a total of 114 visits. For the test set there were a total of 25 normal birth weight newborns, for which 22 had one visit and three had two visits, giving a total of 28 visits. Table 7.5 shows the number of visits for each GA for the training and test sets.

Table 7.5: Number of visits per gestational age (GA) for the 104 normal birth weight training set, the 25 normal birth weight test set, and the 24 low birth weight test set.

Gestational age (months)	Training set	Test set	Test set
	Normal birth weight	Normal birth weight	Low birth weight
6	6	3	3
7	28	10	7
8	32	5	5
9	48	10	10
total	114	28	25

Training/Validation methodology

The training and validation procedure was performed using a five fold-cross validation with 50 trials (repetitions). At each trial, the patients were randomly assigned to different folds, ensuring that visit features corresponding to the same patient were in the same folder. By using 50 trials the variability of the models for estimating the GA could be measured, and confidence intervals could be estimated.

Since folds were class unbalanced (see Table 7.2), at each iteration of the five-fold cross-validation, the number of visits per GA was balanced on the training folds before constructing a model. To that end, we used the Adaptive Synthetic Sampling (ADASYN) method, which has been reported to overcome the class imbalance problem in support vector machine models [36]. This method generates synthetic data for the minority classes by taking the Euclidean distance between two data points and then adding the difference scaled by a factor, between 0 and 1, to one of the minority data points. In this study, the ADASYN was implemented as described in [147].

Before training a model, the balanced training set and the held-out fold set were standardized by subtracting the mean of the respective feature vector and dividing it by its standard deviation computed in the training data only. This standardization method was selected as it has shown to be suitable for feature scaling in machine learning methods [359].

The 50-trial five-fold cross-validation was assessed using three different regression methods performed to assess three regression approaches: Elastic Net, Support Vector

Regression (SVR), and Gradient Boosting Tree (GBT). At each iteration of the cross-validation, the training folds were used to select the most relevant features for the SVR and GBT models. For the SVR, features were selected using the maximum relevance and minimum redundancy (mRMR) algorithm [285], which ranks the most relevant features based on mutual information gain. For the GBT model, the features were selected by training a GBT with 100 trees and learning rate of 1 on the training folds, and then identifying the most relevant features by summing the feature weights over all the weak learners.

To optimize each model's hyperparameters, a nested cross-validation using a grid search on the training folds was used for each model. The grid search for the three models was defined as:

- Elastic Net. The linear penalty term, λ was defined as $\{0.1, 0.2, \dots, 0.8, 0.9\}$. The quadratic penalty term was given by $\frac{1-\lambda}{2}$. For each λ value, a set of 100 values of regularization parameters were tested. The regularization parameter set was generated by first finding the largest value, θ , that gave a non-null model (i.e. intercept $\neq 0$), and then the remaining 99 values were defined by decreasing θ by 10^{-5} , so that the ratio of the smallest to the largest value of the set was 10^{-4} .
- SVR. The grid search for the soft margin (C) and the margin of tolerance (ϵ) were defined as: $C \in \{2^{-3}, 2^{-1}, \dots, 2^8\}$ and $\epsilon \in \{2^{-10}, 2^{-9}, \dots, 2^{-5}\}$. The Gaussian radial basis function parameter (γ) was analytically estimated as reported in [61]. Namely, γ was derived by calculating the distribution of $\|x - x'\|^2$ between a subset containing 70% of the training set, and then taking the inverse of the median of this distribution.
- GBT. The grid search for the learning rate was defined as $\{0.1, 0.25, 0.5, 1\}$ and the number of trees was defined as $\{100, 150, 200, \dots, 500, 550\}$. The number of

maximum splits (tree height) was defined as $\{1, 2, \dots, \log_2(S - 1)\}$; where S is the total number of visits of the training folds.

Analyzing the Training/Validation output

The 50 trial five-fold cross-validation resulted in 50 median absolute error (MAE) vectors, and 250 selected feature vectors. From the 50 MAE vectors, the median, interquartile range, and the lower and upper 95% confidence interval for the median were determined. The median MAEs were compared to select which regression model (Elastic Net, SVR, or GBT) to use in the test stage.

From the 250 selected feature vectors, the top twenty most relevant features were identified. To that end, the features were ranked at each feature vectors, and the mean rank was determined by averaging the ranking of each feature over the 250 feature vectors. This simple aggregation technique was used as it has shown to be effective to combine different features sets in the medical application field [313, 406].

Using the top ten, top fifteen, and top twenty, the same validation/training procedure was repeated to identify the best performing feature set for estimating GA.

Testing methodology

The model selected with the best performing feature set were then used to train a final model on the training set of normal birth weight patients. Before training the final model, the training data were balanced using ADASYN [147], and the parameters were optimized using grid search as explained in subsection 7.4.4. The final model was then used to estimate the GA for both the 25 held-out normal birth weight and the 24 LBW patients.

Since the final model depends on the nature of the synthetic data added to the training dataset, the testing procedure was performed 100 times to evaluate the variability of the model's performance. The median, interquartile range, and the lower

and upper 95% confidence interval for the median were determined for the two test groups. To determine if there was any difference in GA estimation distribution of errors between the normal and LBW newborns, a two-sided Wilcoxon rank-sum test hypotheses test was evaluated on the data.

Univariate sources

In order to assess the relevance of using a combination of 1D-DUS and maternal blood pressure based features, the GA estimation procedure was repeated using features from each source of data separately. To that end, the methods presented in subsections 7.4.4, 7.4.4, and 7.4.4 were repeated for features extracted for each source.

In a manner similar to the training/validation procedure, the best feature set was found for each monitoring modality. From the 27 1D-DUS based features, the top twenty, fifteen, and ten features were identified in the training set. From the ten maternal blood pressure-based features, the top ten, eight, and five features were identified. Although an exhaustive search may have identified a better performing set of features, we considered this process a reasonable trade off between overfitting and parsimony.

The 100 MAE obtained using only features extracted from 1D-DUS and the 100 MAE obtained using only features derived from maternal blood pressure were compared to the 100 MAE obtained using both types of features. The comparison was performed using a one-sided Wilcoxon rank-sum test with the alternative hypothesis that the median MAE of the individual source feature is lower than the median MAE of combining features derived from both sources.

7.4.5 Detecting possible cases of IUGR

Since we are assuming that IUGR cases are those with LBW, the estimated GA for the test set were compared against birth weight. To that end, the GA error estimation

was defined as the difference between the GA based on the LMP and the median GA estimation over the 100 repetition. Then, a robust least square was fitted using the birth weight as independent variable and the GA error estimation as the response variable.

7.5 Results

7.5.1 Training/Validation performance

Table 7.6 shows the mean absolute error (MAE) for the 50 trial five-fold cross-validation. For all the three regression models, the MAE of the seventh and eighth gestational months were lower than those of the extreme months evaluated. The regression model with the lowest the overall median MAE over the 50 trials was the SVR with a value of 0.8 months. Furthermore, the SVR and Elastic Net were the models with the lowest interquartile range for the MAE over the 50 trials, thus indicating a lower variance of these to models in comparison to the GBT.

Table 7.6: Mean absolute errors (MAE) of the 50 trial five-fold cross validation for the Elastic Net, SVR, and Gradient boosting tree. For each model, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided.

Model	Metric	Gestational age (months)				
		6	7	8	9	All
Elastic Net	Median	1.54	0.51	0.50	1.40	0.93
	IQR	0.10	0.04	0.07	0.14	0.08
	LCI	1.52	0.50	0.49	1.36	0.91
	UCI	1.58	0.52	0.52	1.42	0.96
SVR	Median	1.57	0.51	0.43	1.15	0.80
	IQR	0.17	0.09	0.10	0.11	0.06
	LCI	1.54	0.48	0.40	1.11	0.79
	UCI	1.65	0.54	0.47	1.17	0.83
Gradient boosting tree	Median	1.77	0.85	0.65	0.81	0.86
	IQR	0.42	0.20	0.14	0.18	0.08
	LCI	1.66	0.82	0.61	0.77	0.83
	UCI	1.93	0.89	0.70	0.89	0.88

7.5.2 Ranking the features

Table 7.7 shows the top twenty features for estimating GA based on the average ranking of the 250 feature vectors for the SVR. Seven out of the ten top features were derived from the 1D-Doppler ultrasound, being fHRV linear indexes the most common. Maternal blood pressure based features were also included in the top features, being the MAP the most important feature of that group.

Table 7.7: Feature ranking obtained after averaging the individual 250 feature ranking resulted in the 50 trial five-fold cross-validation.

Ranking	Feature	Type
Top10	Tone	Non-linear and Complexity
	#acc/min	Linear
	MAP	Hemodynamic formula
	ApEn	Non-linear and Complexity
	KurtosisIS	Linear
	CO	Hemodynamic formula
	STV/LTV	Linear
	SkewnessIS	Linear
	PP	Hemodynamic formula
	LTI	Linear
Top15	stdIS/rmssdIS	Linear
	DBP	Hemodynamic formula
	RPP	Hemodynamic formula
	varIS	Linear
	stdIS	Linear
Top20	GMI	Non-linear and Complexity
	II	Linear
	DAC	Linear
	MF	Spectral
	basal FHR	Linear

Table 7.8 presents the results obtained by repeating the training/validation procedure using the top10, top15, and top20 features. The top fifteen feature set achieved the lowest median MAE over the 50 trials with a value of 0.76 months (95%CI = 0.75-0.78 months).

Table 7.8: Mean absolute errors (MAE) of the 50 trial five-fold cross validation for the SVR using the top10, top15, and top20 features. For each model, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided.

Model	Metric	Gestational age (months)				
		6	7	8	9	All
SVR (top ten features)	Median	1.62	0.48	0.43	1.18	0.82
	IQR	0.18	0.10	0.09	0.12	0.07
	LCI	1.57	0.45	0.40	1.13	0.80
	UCI	1.67	0.50	0.45	1.21	0.83
SVR (top fifteen features)	Median	1.51	0.47	0.44	1.04	0.76
	IQR	0.17	0.08	0.07	0.15	0.07
	LCI	1.44	0.45	0.42	1.01	0.75
	UCI	1.54	0.50	0.46	1.08	0.78
SVR (top twenty features)	Median	1.56	0.45	0.43	1.16	0.81
	IQR	0.15	0.10	0.08	0.09	0.05
	LCI	1.52	0.42	0.40	1.14	0.79
	UCI	1.60	0.47	0.44	1.20	0.81

7.5.3 Testing performance

Table 7.9 shows the performance of the 100 repetitions training an SVR with the top15 on the training set and testing on the held-out 25 normal birth weight newborns and the 24 LBW newborns. The median MAE for each gestational month and the overall was statistically significantly higher for the LBW newborns (two-sided Wilcoxon rank-sum test; $\alpha = 0.05$). The difference between the median MAE for the two groups was increasing throughout the GA, resulting in a difference of 0.29 months for the overall estimation.

Figure 7.2 shows the difference (δ) between GA based on the LMP and the median estimated GA over the 100 repetitions for each visit. The LBW newborns' (red crosses) GAs were generally overestimated (LMP-GA < estimated GA) for the eighth and ninth gestational months compared to the normal birth weight newborns (blue circles). For the eighth and ninth gestational months, on the other hand, the LBW newborns were generally underestimated (LMP-GA > estimated GA) compared to the normal birth weight newborns.

Table 7.9: Mean absolute errors (MAE) of the 100 trials on the test (held-out) normal birth weight and LBW newborns. For each type of newborns, the median, interquartile range, and the 95% confidence interval for the median of the MAE are provided. A † indicates a significant difference between the median GA estimations of the normal and LBW newborns for the 100 repetitions (two-sided Wilcoxon rank-sum test; $\alpha = 0.05$)

Newborn type	Metric	Gestational age (months)				
		6 †	7 †	8 †	9 †	All †
Normal birth weight	Median	1.06	0.53	0.33	0.99	0.72
	IQR	0.25	0.05	0.07	0.07	0.05
	LCI	1.03	0.52	0.32	0.98	0.71
	UCI	1.08	0.54	0.35	1.00	0.72
Low birth weight	Median	1.26	0.73	0.68	1.32	1.01
	IQR	0.11	0.05	0.09	0.06	0.03
	LCI	1.24	0.72	0.67	1.30	1.01
	UCI	1.29	0.74	0.69	1.33	1.02

For newborns with more than one visit, Figure 7.2 shows a line connecting the median error across GA. For both normal and LBW newborns, the GA estimates trended from overestimation to underestimations as GA increased. However, the discrepancy was higher for the LBW newborn with a difference of around 2.5 months between the seventh and the ninth gestational month. In contrast, the maximum difference for normal birth weight newborns was approximately 0.75 months.

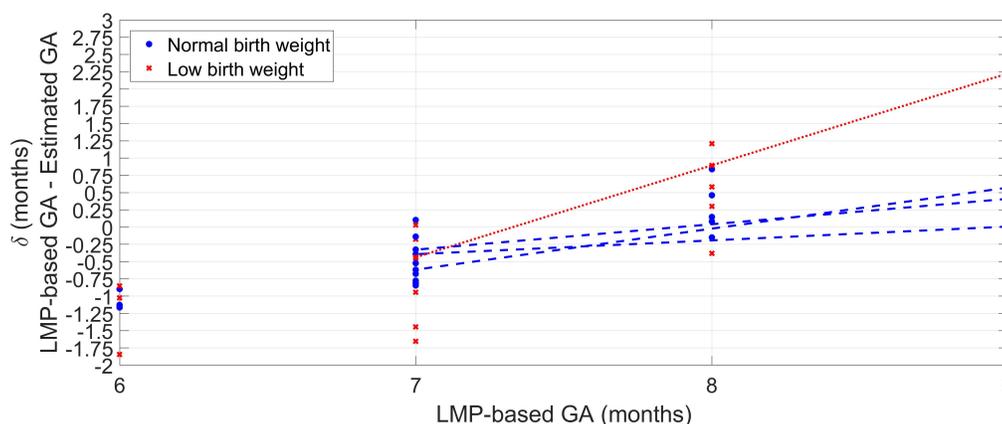


Figure 7.2: Median error of the 100 repetitions against GA provided by the LMP method for the normal weight newborns (blue circles) and the LBW newborns (red crosses). For newborns with more than one visit, a line connects the median error along with the GA.

Figure 7.3 shows the median of the estimated GA for each label of the LMP

method. For all the gestational months, the LBW group resulted in a lower median GA estimations than the normal birth weight group. The median difference between normal and LBW newborns is greater for the last two months of pregnancy, thus indicating a higher inconsistency in the features for GA estimation between the type of newborns from the eighth month of gestation onward.

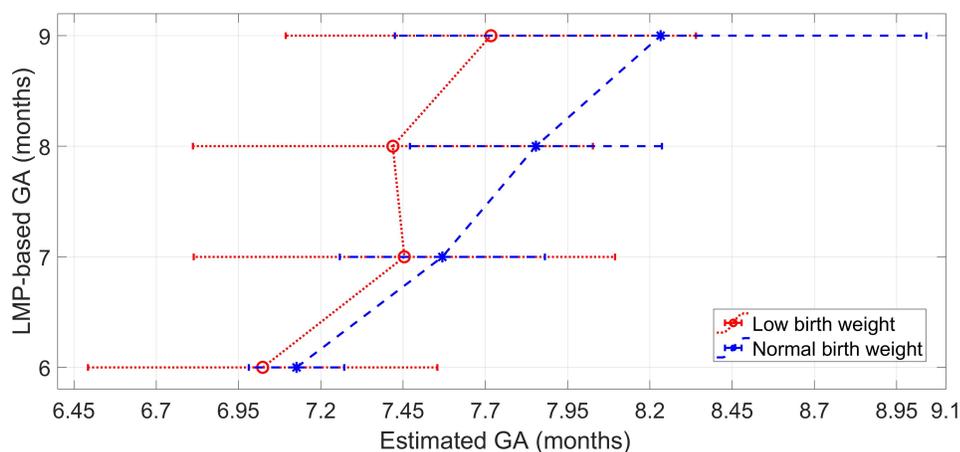


Figure 7.3: Median and interquartile range of the estimated GA for each label of the LMP method for the normal weight newborns and the LBW newborns.

7.5.4 Comparing features for the estimation of GA

For the features derived from the 1D-DUS signals, the lowest validation median MAE was achieved with a SVR using the DUS top ten features: tone, number of acceleration per minute, Approximate entropy, STV/LTV ratio, SkewnessIS, stdIS/rmsdIS ratio, KurtosisIS, LTI, varIS, and II. For the features derived from maternal blood pressure, the lowest validation median MAE was achieved with a SVR using the maternal blood pressure top eight features: PP, MAP, CO, DBP, RPP, SBP, and MHR.

Table 7.10 shows the median MAE for each source of features on the test data for normal and low weight newborns. For the normal birth weight newborns, the features derived from 1D-DUS signals obtained a statically significantly lower median MAE for the seventh and eighth gestational month than using features derived from both

sources. Using features based on maternal blood pressure, on the other hand, were able to obtain significant lower MAE for the sixth gestational month. However, any of the individual source features provided lower median MAEs for the ninth and overall GA estimations.

Table 7.10: Median of the mean absolute errors (MAE) of the 100 trials on the test (held-out) normal birth weight and LBW newborns using separately the one-dimension Doppler ultrasound (1D-DUS) based features and maternal blood pressure (MBP) based features. The median MAE of each source of feature were compared to the MAE of combining both sources shown in Table 7.9. The comparison was performed using a one-sided Wilcoxon rank-sum test with the alternative hypothesis that the median MAE of the individual feature source is lower than the median MAE of combining features of both sources. * indicates a statistically significant difference regarding (left one-sided Wilcoxon rank-sum test; $\alpha = 0.05$)

Newborn type	DUS-based features (SVR with top10)					MBP-based features (SVR with top8)				
	Gestational age (months)					Gestational age (months)				
	6	7	8	9	All	6	7	8	9	All
Normal birth weight	1.42	0.51*	0.26*	1.18	0.80	0.87*	0.78	0.79	1.15	0.92
Low birth weight	1.25*	0.52*	0.47*	1.38	0.95*	1.02*	0.65*	0.53*	1.64	1.07

For LBW patients, both sources of features (DUS and BP) obtained lower significant MAEs for the sixth, seventh, and eighth gestational months. The features derived from 1D-DUS signals led to a statically significantly lower median MAE for the overall GA estimations.

7.5.5 GA estimation errors as a function of birth weight

Figure 7.4 shows the GA estimation errors over estimated birth weight for the tested newborns. Robust least-square fits were performed for each type of newborn, as well as for all the newborns as a whole. All the fits provided negative slopes and negative Pearson correlation values (ρ). The inverse relationship between GA estimation error and birth weight indicates that there are more underestimations for newborns with LBW. In fact, for the LBW newborns, fifteen visits achieved underestimations, whereas ten visits obtained overestimations. For the normal weight newborns, the fitted line was $\delta_{NBW} = 1.85 - 0.60w_{NBW}$ ($\rho = -0.15$, P -value = 0.45). For the low

weight newborns, the fitted line was $\delta_{LBW} = 0.82 - 0.23w_{LBW}$ ($\rho = -0.06$, P -value = 0.76). For all the newborns, the fitted line was $\delta = 1.12 - 0.35w$ ($\rho = -0.13$, P -value = 0.37).

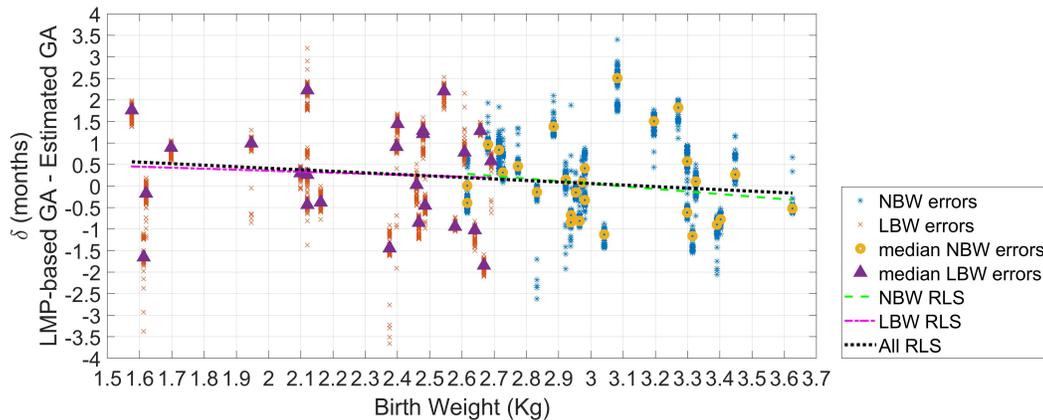


Figure 7.4: Error in GA from 100 repetitions (δ) as the difference between GA based on the LMP and the estimated GA. The median error of the 100 repetitions for each recording (or visit) is displayed (triangles for LBW newborns; and circles for normal birth weight newborns). Robust least square (RLS) fits are also shown.

7.6 Discussion

7.6.1 Interpretations of Findings

The results presented in this work indicate that it is possible to provide a proxy for fetal growth assessment in a resource-constrained setting by using the difference between GA estimated by LMP and the GA estimated from features extracted from an inexpensive Doppler transducer and a blood pressure device.

This proxy fetal assessment relies on the GA estimation approach introduced in this work, which using a pregnancy conversion factor of $\frac{40 \text{ weeks}}{9 \text{ months}}$ resulted in a median MAE of 3.2 and 4.5 weeks for the normal and LBW newborns, respectively. Interestingly, these MAE values are comparable to those presented in [235] and [237] of 2.7 to 5.1 weeks obtained using a step-wise regression using 1D-DUS and fECG signals recorded by medical professionals in a high-resource / high-income country.

Moreover, unlike [360], our work did not require high-resolution input signals, making the implementation of the approach described here feasible in LMICs. Notably, our GA estimations were lower than that at the six week point reported for the SFH measurement method [139].

The higher GA estimation errors for the LBW newborns (see Table 7.9) indicate that this type of patient has different patterns in the 1D-DUS and maternal blood pressure features than normal birth weight newborns of the same GA. We note that this difference is related to the birth weight (Figure 7.4), and therefore assuming that LBW is a consequence of IUGR [205], a potential sign of IUGR can be detected when the estimated GA is lower than the GA calculated from last menstruation period. This provides evidence to indicate that our method is a low-cost alternative fetal growth assessment to identify cases that need to be referred to further medical assistance in LMICs, in which SFH measurement is not sufficiently accurate [403], and ultrasound imaging is not available [399].

The longitudinal changes in the difference between GA estimations of low and normal birth weight newborns across gestation suggests that IUGR is progressive and is more evident for the eighth and ninth gestational months (see Figure 7.3). Therefore, our proxy method may be more effective in detecting fetal growth abnormalities during the last two months of gestation, thus helping to identify fetuses that need assistance during delivery to reduce adverse perinatal outcomes.

Another interesting finding was the selected features for estimating GA. The top15 features (Table 7.7) were consistent with features previous work for assessing gestational development [380, 379, 332, 386, 198, 185]. Specifically, linear, non-linear and complexity features, such as tone, the number of accelerations per minute, approximate entropy, and statistical movements of the interbeat sequence, were the features which provided the SVR with the highest performance boost. The feature selection algorithm also demonstrated the potential of blood-pressure-derived features. This

selection was relevant as little research has used this type of features for assessing fetal maturity. Finally, the STV/LTV ratio, which previously has shown to be relevant for detecting IUGR cases [355], was also relevant for GA estimation.

Notably, the combination of features from both sources (Doppler and blood pressure) resulted in a lower MAE than estimating GA using features from only one source. Based on the individual performance of 1D-DUS and maternal blood pressure based features (Table 7.10), the maternal blood pressure-based features help to reduce estimations errors in the sixth gestational month, whereas the 1D-DUS based features contributed to reducing estimation errors in the seventh and eight gestational months. Therefore, the inclusion of both source of features seems relevant to reduce GA estimation error, as well as identify possible cases of LBW.

7.6.2 Study Limitations

It should be noted that in this work possible cases of IGUR were defined by newborns birth weight. This assumption could not be validated as patients did not receive a ultrasound imaging exam to detect IUGR based on obstetrician standards. Nevertheless, this assumption was based on the fact that in LMICs around 60% of the LBW is caused by IUGR [205].

The method presented here estimates GA using the LMP method as a reference. As LMP is not a completely unbiased method for dating fetuses [20, 95], our results may contain a bias. Moreover, since the errors in GA estimation provided by our method were larger than two weeks (the estimation error recommended in the literature [218]), our method is not accurate enough to be used as a primary method for dating GA. However, for rural areas in LMICS, in which there is a lack of ultrasound imaging equipment and obstetricians, our method is a proxy to detect fetuses with possible abnormalities (LBW or IUGR) that need to be referred for further medical diagnosis and treatment. Superiority to the SFH measurement indicates that this

method should be preferred.

Another limitation of this work was that the GA was recorded by the clinical team in months rather than weeks [232]. However, as a month includes a variable number of days, this introduces a quantization/rounding error - fetuses just a few days apart that fall into different months will look similar but will be identified as different. This decreases the accuracy of any model fitted to the data, resulting in larger absolute errors for the sixth and ninth gestational months (Table 7.9).

This error can be thought of as a higher intra-class variance. When intra-class variance is high, it is recommended to use a longitudinal approach rather than cross-sectional one [96]. However, those models need multiple points per subject, in order to be able to apply mixed models considering the random effects of each individual. In this study we could not apply a longitudinal approach because the majority of subjects contained only one valid visit. Nevertheless, the MAE values obtained for our approach suggests that features and methods used here are promising for estimating GA based on the LMP method, which is a low-cost, feasible method to date pregnancy in LMICs [265].

Our study also included visits that were between the sixth and ninth months of gestation. To fully assess the capacity of our approach to estimating GA, it should be evaluate on metrics recorded in the first and second trimesters. Such an evaluation would allow for the comparison of our GA estimation in a fetal development period in which genetic and biological variability of fetal size is low, and in which Doppler images methods estimate GA more accurately [298].

Finally, we note that the approach presented here did not consider fetal sex to estimate GA. Although it may influence fHRV metrics used here for GA estimation, we deliberately avoid gender because the aim is to avoid the use of imaging Doppler, and sex determination, which present significant cost and social problems respectively.

7.6.3 Future directions

Future research should focus on increasing the temporal resolution of the GA labels (by recording the week of the LMP through community surveys perhaps), and use a more accurate dating method such as expert-driven ultrasound imaging.

Future research should also evaluate the efficacy of the proxy presented here on confirmed diagnoses of IUGR. This evaluation would allow a full end-to-end assessment of how 1D-DUS and maternal blood pressure can contribute to detection fetal growth abnormalities.

Moreover, future research should ensure the collection of multiple visits during the course of pregnancy (and extending this to earlier gestational periods), so that a longitudinal analysis can be performed that incorporates the individual dynamics into the model.

7.7 Conclusion

This work introduced a proxy to detect possible cases of fetal growth retardation for constrained-resource environments in which ultrasound imaging is not available, and current low-cost methods are prone to error. The potential IUGR cases are detected by comparing GA based on the last menstrual period with estimates obtained 1D-DUS and maternal blood pressure recordings collected with inexpensive devices, usable with little training. The method is valuable to endow non-medically trained operators with an objective metric to identify cases that need to be referred to further medical assistance. The assessment system may, therefore, have an immediate impact if coupled with suitable intervention, such as nutritional supplementation. However, a prospective clinical trial is required to show the efficacy of such metrics and intervention.

Chapter 8

Conclusion

8.1 Summary and contributions

The work presented in this thesis addressed the issue of tracking gestational development using a low-cost monitoring system appropriate for LMICs. Specifically, the monitoring system collects 1D-Doppler ultrasound signals (1D-DUS) of fetuses by means of a mobile health system previously introduced by Stroux *et al.* [351, 355] in a Guatemalan highland rural community. Gestational development tracking is a crucial issue in rural Guatemala, where perinatal and neonatal death rates are alarmingly high [368]. A leading contributor to this burden is IUGR [205], which is a slowdown of gestational development leading to low birth weight. Low birthweight, in turn, is associated with lower neurodevelopmental scores and other adverse sequelae [258, 207]. Critical cases of restricted-growth fetuses can be identified by developing methods to assist Traditional Birth Attendants (TBAs), thus allowing for timely referrals and interventions.

Since 1D-DUS are a primary source for the gestational tracking methods presented in this thesis, the first relevant step was to ensure the quality of such collected signals. Chapters 3 and 4 presented methods to assess the quality of 1D-DUS recorded with

the inexpensive Doppler transducer used in the mHealth system. To that end, time, spectral and non-linear features were extracted, and a supervised machine learning approach was used to classify the 1D-DUS into five quality classes observed in the first release cycles of the mHealth app. Taking into account computational time, a final set of 17 features achieved an F1-score higher than 90% to classify the different quality classes. Notably, the used feature set only requires approximately 360 ms to be extracted from a 3.75-second 1D-DUS segment using a smartphone processor, thus enabling the quality assessment functionality to be transferred to the mHealth app.

Functionality to track fetal gestational development was developed using FHR estimations. Chapter 7 described how linear, spectrum and complex features extracted from FHR time-series were combined with maternal blood pressure features to estimate gestational age. The method achieved a mean estimation error of 0.8 months, which is comparable to those provided by previous works [235, 237] performed using CTG and fECG recorded in a hospital environment. Therefore, the method provides a proxy for assessing fetus development. For instance, a potential sign of IUGR can be identified if the estimated GA is lower than the GA reported by the last menstruation. The method also provides an interesting finding regarding the most relevant features for gestational age estimation. In particular, maternal blood-pressure-derived features were shown to be relevant; however, before this work, little research had used them to assess fetus maturity.

Chapter 6 presented a gestational development tracking functionality, aimed at postpartum visits. Given the high prevalence of low birth weight and delayed first postnatal visit in LMICs, a regression model was developed to identify possible low-birth-weight newborns based on observed postnatal weights up to 60 days after birth. Birth weight estimations were consistent with figures reported by the Guatemalan government for the region of study [254], thereby showing the proposed regression model's potential to estimate birth weight. It may be possible to identify LBW new-

borns by weighing infants several days after birth, and, therefore, initiate treatment to reduce short- and long-term complications.

In addition to contributing to fetal gestation tracking functionalities, this thesis contributed to annotate datasets. Specifically, two 1D-DUS datasets and one maternal blood pressure dataset were annotated. The annotation process made all the methods presented here possible. Moreover, the annotated datasets have sparked additional related projects that seek to improve the mHealth system for providing fetal monitoring to populations in LMICs.

8.2 Limitations

The work conducted has some limitations. The methods presented in Chapters 3 to 7 were developed and tested offline. The methods were not tested online at the point of care because this work focused on the development of the functionalities rather than in the implementation of them. Nevertheless, the obtained results were designed and developed taking computational cost into consideration, thereby ensuring the use of inexpensive computational methods that can be timely executed in a mobile application.

Moreover, the functionalities presented in this work are in a prototype stage and have not yet been implemented in the mobile application. Although computational cost was considered as a factor in the development, it will be necessary to transfer the functionalities to a smartphone to assess the real contribution to perinatal care in Guatemala rural communities.

Another limitation regarding gestational tracking functionalities (Chapters 6 and 7) is that the methods were developed and tested with 1D-DUS and maternal blood pressure recordings collected from 2016 to 2018 in rural Chimaltenango. The gestational age and birth weight estimations have not been tested with recordings collected

after 2018 in Chimaltenango, nor with data collected from different LMICs around the world. Therefore, the stability and accuracy of the features used for gestational age and birth weight may not be generalizable to other populations. However, the approach presented here can be adjusted to new LMICS populations, thereby extending the functionalities here to new populations.

This thesis aimed to develop functionalities to track fetal gestational development. However, it should be noted that in this thesis possible cases of IGUR were defined as those newborns whose birth weight was low. This assumption could not be validated as patients did not receive a Doppler imaging exam to detect IUGR based on obstetrician standards.

Similarly, gestational age estimation regression was trained using the last menstrual period as a standard. However, as previously reported in the literature, the last menstrual period is not totally accurate, thereby introducing noise in the development of the method presented in Chapter 7. Again, as Doppler imaging was not available in the Guatemalan rural community, the gestational age labels were not validated. The accuracy of the presented method may, therefore, be affected when tested with a dataset labeled with different gestational age methods. Nevertheless, the approach presented in Chapter 7 can be easily adjusted to a dataset with more accurate gestational age labels such as those provided by Doppler imaging.

8.3 Future work

8.3.1 Online functionalities

All functionalities presented here were developed for offline processing. As the final goal of the mHealth system is to function as a decision support system able to provide timely feedback to TBAs, it is necessary that all functionalities be processed online. To that end, two options are possible. The first option is to implement all

functionalities on the mobile application; thus, enabling the smartphone to process all the 1D-DUS and maternal blood pressure recordings. The implementation of the functionalities is feasible since low-cost computational techniques were used in the development. The second option is to send the recorded data to be processed in a remote server using a cellphone communication network. This option requires a cellphone network plan with enough bandwidth to send and receive information. However, in rural areas of LMICs, the signal is weak as geographical conditions interrupt the transmission between smartphones and cellular antennas. Therefore, the first option seems to be more viable than the second one since, in that case, the functionalities would rely only on the smartphone and not on external factors, such as the cellphone network.

8.3.2 Segmentation of 1D-Doppler ultrasound signals into beat-to-beat intervals

Commercial hand-held and CTG devices use autocorrelation-based methods such as that presented in Chapter 4 for estimating FHR from a 1D-DUS signal. However, previous works have shown that AC-based methods are affected by the inherent smoothing or averaging of the autocorrelation [176, 201, 65]. This averaging characteristic tends to limit the capacity of AC-based approaches to follow rapid FHR decelerations. Indeed, by comparing the performance of AC-based methods to indirect fECG, Jezewski *et al.* [176] found statistically significant differences in fetal heart rate variability indicators, such as decelerations and short term variability.

In order to overcome the problems of AC-based approaches, a 1D-DUS signal should be segmented into beat-to-beat intervals. By segmenting the signal into intervals, accurate estimations of heart rate variability metrics can be performed, similar to those given by fECG. One approach to this was presented in Stroux *et al.* [351], in which a Hidden Markov model-based (HMM) approach was used to segment beats

in a 1D-DUS signal, achieving F1-scores of 91.8% and 98.4% to detect beat-to-beat intervals in 1D-DUS signals of intermediate and good quality signals, respectively. However, the HMM model needed annotations of the heartbeat locations to optimize parameters. Such annotation is time-consuming and subject to high variability among annotators. Future research should address this problem by combining unsupervised methods with the models that improve the segmentation of beats, such as hidden Markov models [347], thereby increasing the accuracy of HRV estimates.

8.3.3 Deep learning for improving gestational age estimations and detection of IUGR cases

This thesis used classical machine learning supervised methods, particularly an SVM, to estimate gestational age. Modern methods of Deep Learning (DL) were not used as the dataset contained fewer than 200 patients (see Chapter 7). However, DL may help to improve estimations of gestational age, as well as of other parameters, such as IUGR. The advantage of DL is that the neural network structure facilitates the identification of features beyond those hand-crafted by the programmer. In the case of 1D-DUS signals, DL algorithms can learn from the raw data without using any previous feature engineering process to determine which positions of a 1D-DUS signal are relevant for the estimations. As the frequency and shape of a 1D-DUS signal depend on the maturity and health of the fetus, DL may be able to associate gestational age and health status labels with the input 1D-DUS signals. Therefore, future work should evaluate the performance of DL in gestational development tracking on larger datasets.

Appendix A

Table of all feature combinations for quality assessment

Table A.1: Median classification performance of the 100 five-fold cross validation balanced with bootstrapping for all the possible feature combinations. The table is grouped for feature vectors of the same length. For each combination of features, the median and interquartile range of the accuracy rate of the 100 repetitions are shown.

Feature Combination	Median Accuracy	Interquartile Range
H_s	<u>84.17%</u>	5.83%
PSD	84.17%	6.67%
SQI4	64.17%	11.25%
SQI3	67.92%	5.00%
SQI2	74.17%	6.67%
SQI1	61.67%	10.42%
PSD, H_s	85.06%	5.00%
SQI4, H_s	84.58%	5.00%
SQI4,PSD	78.42%	15.42%
SQI3, H_s	83.33%	6.67%
SQI3,PSD	84.17%	12.92%

Continue in the next page

Table A.1 – Continuation of the previous page

Feature Combination	Median Accuracy	Interquartile Range
SQI3,SQI4	65.00%	9.58%
SQI2, H_s	<u>85.83%</u>	5.00%
SQI2,PSD	84.65%	10.83%
SQI2,SQI4	69.17%	5.83%
SQI2,SQI3	73.33%	7.50%
SQI1, H_s	85.83%	5.46%
SQI1,PSD	81.67%	8.33%
SQI1,SQI4	61.67%	9.17%
SQI1,SQI3	62.50%	7.22%
SQI1,SQI2	71.67%	8.75%
SQI4,PSD, H_s	81.67%	8.75%
SQI3,PSD, H_s	85.00%	7.50%
SQI3,SQI4, H_s	81.67%	5.83%
SQI3,SQI4,PSD	80.83%	10.83%
SQI2,PSD, H_s	<u>85.83%</u>	5.00%
SQI2,SQI4, H_s	84.17%	5.00%
SQI2,SQI4,PSD	83.33%	8.33%
SQI2,SQI3, H_s	84.17%	5.47%
SQI2,SQI3,PSD	80.00%	9.58%
SQI2,SQI3,SQI4	67.50%	8.33%
SQI1,PSD, H_s	85.00%	6.67%
SQI1,SQI4, H_s	82.50%	6.25%
SQI1,SQI4,PSD	78.33%	12.92%
SQI1,SQI3, H_s	83.33%	5.00%
SQI1,SQI3,PSD	82.50%	14.58%
SQI1,SQI3,SQI4	65.00%	7.08%

Continue in the next page

Table A.1 – Continuation of the previous page

Feature Combination	Median Accuracy	Interquartile Range
SQI1,SQI2, H_s	85.00%	5.83%
SQI1,SQI2,PSD	84.17%	8.33%
SQI1,SQI2,SQI4	69.17%	5.83%
SQI1,SQI2,SQI3	72.50%	6.67%
SQI3,SQI4,PSD, H_s	80.83%	7.50%
SQI2,SQI4,PSD, H_s	<u>85.00%</u>	8.33%
SQI2,SQI3,PSD, H_s	82.92%	5.83%
SQI2,SQI3,SQI4, H_s	83.33%	5.83%
SQI2,SQI3,SQI4,PSD	77.08%	9.17%
SQI1,SQI4,PSD, H_s	82.50%	7.83%
SQI1,SQI3,PSD, H_s	85.00%	10.83%
SQI1,SQI3,SQI4, H_s	79.17%	5.00%
SQI1,SQI3,SQI4,PSD	79.17%	12.50%
SQI1,SQI2,PSD, H_s	84.17%	6.25%
SQI1,SQI2,SQI4, H_s	82.50%	5.83%
SQI1,SQI2,SQI4,PSD	80.83%	7.15%
SQI1,SQI2,SQI3, H_s	82.92%	6.67%
SQI1,SQI2,SQI3,PSD	79.17%	8.75%
SQI1,SQI2,SQI3,SQI4	68.33%	7.08%
SQI2,SQI3,SQI4,PSD, H_s	80.83%	10.00%
SQI1,SQI3,SQI4,PSD, H_s	77.50%	10.83%
SQI1,SQI2,SQI4,PSD, H_s	<u>84.65%</u>	5.00%
SQI1,SQI2,SQI3,PSD, H_s	84.17%	6.67%
SQI1,SQI2,SQI3,SQI4, H_s	82.50%	5.00%
SQI1,SQI2,SQI3,SQI4,PSD	77.50%	8.33%
SQI1,SQI2,SQI3,SQI4,PSD, H_s	<u>83.75%</u>	6.67%

Appendix B

Algorithm for determining periodicity

FindingPeriodicity Algorithm for determining periodicity of 1D-DUS signal

procedure FINDINGPERIODICITY

peaks \leftarrow Peak amplitude of the AC window, sorted by location

locs \leftarrow Sorted location of the AC window peaks

main:

if $length(peaks) > 1$ **then** \triangleright Are there more than 1 pronounced peaks?

ratio $\leftarrow locs[1]/locs[2]$ \triangleright Finding ratio between peak times

if $ratio \geq 0.48$ and $ratio \leq 0.52$ **then**

 return *loc*[1] \triangleright In this case, peaks are harmonic

else

if $peaks[1]/peaks[2] > threshold$ **then**

 return *loc*[1]

else

 return *loc*[2]

else

 return *loc*[1]

Bibliography

- [1] S. Abe. *Support vector machines for pattern classification*, volume 2. Springer, 2005.
- [2] A. Z. Abuhamad and R. Chaoui. *A practical guide to fetal echocardiography: normal and abnormal hearts*. Lippincott Williams & Wilkins, 2012.
- [3] 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.
- [4] J. Agnew. The causes and effects of distortion and internal noise in hearing aids. *Trends in Amplification*, 3(3):82–118, 1998.
- [5] W. Aherne and M. Dunnill. Quantitative aspects of placental structure. *The Journal of Pathology and Bacteriology*, 91(1):123–139, 1966.
- [6] H. Al-Angari, Y. Kimura, L. Hadjileontiadis, and A. Khandoker. A hybrid EMD-kurtosis method for estimating fetal heart rate from continuous Doppler signals. *Frontiers in Physiology*, 8:641, 2017.
- [7] 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.

- [8] 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.
- [9] Z. Alfirevic, T. Stampalija, and T. Dowswell. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database of Systematic Reviews*, (6), 2017.
- [10] L. D. Allan. *Manual of fetal echocardiography*. MTP Press, Lancaster, UK, 1986.
- [11] L. D. Allan. A practical approach to fetal heart scanning. *Seminars in Perinatology*, 24(5):324–330, 2000.
- [12] 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.
- [13] 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.
- [14] I. Amer-Wählin, 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.
- [15] 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.

- [16] American College of Obstetricians. Gestational diabetes. FAQ177, 2013. Retrieved from <http://www.acog.org/Patients/FAQs/Gestational-Diabetes> on Jan 11 2020.
- [17] American College of Obstetricians and Gynecologists. Hypertension in pregnancy. report of the american college of obstetricians and gynecologists' task force on hypertension in pregnancy. *Obstetrics and Gynecology*, 122(5):1122, 2013.
- [18] American Pregnancy Association. Miscarriage, 2019. Retrieved from <https://americanpregnancy.org/pregnancy-complications/miscarriage/> on Jan 11 2020.
- [19] L. M. Anchieta, C. C. Xavier, and E. A. Colosimo. Growth of preterm newborns during the first 12 weeks of life. *Jornal de Pediatria*, 80(4):267–276, 2004.
- [20] 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.
- [21] J. B. Andersen and G. F. Pedersen. The technology of mobile telephone systems relevant for risk assessment. *Radiation Protection Dosimetry*, 72(3-4):249–257, 1997.
- [22] S. Andersson. Acceleration and deceleration detection and baseline estimation. Master's thesis, Chalmerss University of Technology, Göteborg, Sweden, 2011.
- [23] F. Andreotti. *Extraction and detection of fetal electrocardiograms from abdominal recordings*. PhD thesis, Technische Universität Dresden, Leipzig, Germany, 2016.

- [24] T. Arcangeli, B. Thilaganathan, R. Hooper, K. Khan, and A. Bhide. Neurodevelopmental delay in small babies at term: a systematic review. *Ultrasound in Obstetrics and Gynecology*, 40(3):267–275, 2012.
- [25] N. Archer and N. Manning. *Fetal Cardiology*. Oxford University Press, Oxford, UK, 2009.
- [26] F. E. Avni, L. Guibaud, Y. Robert, V. Segers, F. Zierysen, M.-H. Delaet, and T. Metens. MR imaging of fetal sacrococcygeal teratoma: diagnosis and assessment. *American Journal of Roentgenology*, 178(1):179–183, 2002.
- [27] 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.
- [28] D. E. Ayala and R. C. Hermida. Ambulatory blood pressure monitoring for the early identification of hypertension in pregnancy. *Chronobiology International*, 30(1-2):233–259, 2013.
- [29] 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.
- [30] D. Ayres-de Campos, C. Y. Spong, E. Chandrachan, 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.
- [31] P. Bakibinga, E. Kamande, M. Omuya, A. K. Ziraba, and C. Kyobutungi. The role of a decision-support smartphone application in enhancing community

- health volunteers' effectiveness to improve maternal and newborn outcomes in Nairobi, Kenya: quasi-experimental research protocol. *BMJ*, 7(7):e014896, 2017.
- [32] 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.
- [33] R. Bammer. Basic principles of diffusion-weighted imaging. *European Journal of Radiology*, 45(3):169–184, 2003.
- [34] D. J. Barker. Adult consequences of fetal growth restriction. *Clinical Obstetrics and Gynecology*, 49(2):270–283, 2006.
- [35] A. A. Baschat. Ductus venosus Doppler for fetal surveillance in high-risk pregnancies. *Clinical Obstetrics and Gynecology*, 53(4):858–868, 2010.
- [36] R. Batuwita and V. Paladey. Class imbalance learning methods for support vector machines. In H. He and Y. Ma, editors, *Imbalanced Learning*, chapter 5, pages 83–96. IEEE Press Editorial Board 2013, Piscataway, NJ, USA, 1 edition, 2013.
- [37] 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.
- [38] 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.

- [39] J. Behar, J. Oster, Q. Li, and G. D. Clifford. ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Transactions on Biomedical Engineering*, 60(6):1660–1666, June 2013.
- [40] N. A. Bello, J. J. Woolley, K. L. Cleary, L. Falzon, B. S. Alpert, S. Oparil, G. Cutter, R. Wapner, P. Muntner, A. T. Tita, and D. Shimbo. Accuracy of blood pressure measurement devices in pregnancy: a systematic review of validation studies. *Hypertension*, 71(2):326–335, 2018.
- [41] B. R. Benacerraf, T. D. Shipp, B. Bromley, and D. Levine. What does magnetic resonance imaging add to the prenatal sonographic diagnosis of ventriculomegaly? *Journal of Ultrasound in Medicine*, 26(11):1513–1522, 2007.
- [42] 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.
- [43] V. Berghella, L. Pagotto, M. Kaufman, J. C. Huhta, and R. J. Wapner. Accuracy of prenatal diagnosis of congenital heart defects. *Fetal Diagnosis and Therapy*, 16(6):407–412, 2001.
- [44] 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.
- [45] C. S. Berkey. Comparison of two longitudinal growth models for preschool children. *Biometrics*, 38(1):221–234, 1982.
- [46] C. S. Berkey and R. B. Reed. A model for describing normal and abnormal growth in early childhood. *Human Biology*, 59(6):973–987, 1987.

- [47] 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.
- [48] R. E. Black, L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. De Onis, M. Ezzati, C. Mathers, and J. Rivera. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet*, 371(9608):243–260, 2008.
- [49] J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986.
- [50] 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.
- [51] 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.
- [52] C. L. Bose, M. Bauserman, R. L. Goldenberg, S. S. Goudar, E. M. McClure, O. Pasha, W. A. Carlo, A. Garces, J. L. Moore, M. Miodovnik, and M. Koso-Thomas. The global network maternal newborn health registry: a multi-national, community-based registry of pregnancy outcomes. *Reproductive Health*, 12(2):S1, 2015.
- [53] A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis:*

- the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford, Oxford, UK, 1997.
- [54] G. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, Hoboken, NJ, USA, 2015.
- [55] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. LOF: identifying density-based local outliers. In *ACM Sigmod Record*, pages 93–104. ACM, 2000.
- [56] L. Breyssem, H. Bosmans, S. Dymarkowski, D. Van Schoubroeck, I. Witters, J. Deprest, P. Demaerel, D. Vanbeckevoort, C. Vanhole, P. Casaer, and M. Smet. The value of fast MR imaging as an adjunct to ultrasound in prenatal diagnosis. *European Radiology*, 13(7):1538–1548, 2003.
- [57] C. C. Bristow, L. Severe, J. W. Pape, M. Javanbakht, S.-J. Lee, W. S. Comulada, and J. D. Klausner. Dual rapid lateral flow immunoassay fingerstick wholeblood testing for syphilis and hiv infections is acceptable and accurate, port-au-prince, haiti. *BMC infectious diseases*, 16(1):302, 2016.
- [58] H. Brunel, N. Girard, S. Confort-Gouny, A. Viola, K. Chaumoitre, C. D’ercole, D. Figarella-Branger, C. Raybaud, P. Cozzone, and M. Panuel. Fetal brain injury. *Journal of Neuroradiology*, 31(2):123–137, 2004.
- [59] S. Campbell, S. Bewley, and T. Cohen-Overbeek. Investigation of the uteroplacental circulation by doppler ultrasound. In *Seminars in Perinatology*, volume 11, pages 362–368, 1987.
- [60] 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.

- [61] B. Caputo, K. Sim, F. Furesjo, and A. Smola. Appearance-based object recognition using SVMs: which kernel should I use? In *Proceedings of NIPS workshop on Statistical methods for computational experiments in visual processing and computer vision*, volume 2002, Whistler, BC, Canada, 2002.
- [62] 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.
- [63] M. Cassart, A. Massez, T. Metens, F. Rypens, M. A. Lambot, M. Hall, and F. E. Avni. Complementary role of MRI after sonography in assessing bilateral urinary tract anomalies in the fetus. *American Journal of Roentgenology*, 182(3):689–695, 2004.
- [64] Centers for Disease Control and Prevention. PRAMS questionnaires, 2016. Retrieved from <https://www.cdc.gov/prams/questionnaire.htm#questions> on Jun 10 2019.
- [65] M. Cesarelli, M. Romano, and P. Bifulco. Comparison of short term variability indexes in cardiotocographic foetal monitoring. *Computers in Biology and Medicine*, 39(2):106–118, 2009.
- [66] C. C. Chang and C. J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [67] C. J. Chantry, L. A. Nommsen-Rivers, J. M. Peerson, R. J. Cohen, and K. G. Dewey. Excess weight loss in first-born breastfed newborns relates to maternal intrapartum fluid balance. *Pediatrics*, 127(1):e171–e179, 2011.
- [68] 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.
- [69] N. V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1):1–6, 2004.
- [70] 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.
- [71] G. Cicolini, C. Pizzi, E. Palma, M. Bucci, F. Schioppa, A. Mezzetti, and L. Manzoli. Differences in blood pressure by body position (supine, fowler’s, and sitting) in hypertensive subjects. *American Journal of Hypertension*, 24(10):1073–1079, 2011.
- [72] 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.
- [73] G. D. Clifford, F. Azuaje, and P. McSharry. *Advanced methods and tools for ECG data analysis*. Artech House Norwood, Norwood, MA, USA, 2006.
- [74] G. D. Clifford, J. Behar, Q. Li, and I. Rezek. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological Measurement*, 33(9):1419, 2012.
- [75] G. D. Clifford, D. Lopez, Q. Li, and I. Rezek. Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments. In *2011 Computing in Cardiology*, pages 285–288, Hangzhou, China, Sept 2011.

- [76] G. D. Clifford, I. Silva, J. Behar, and G. B. Moody. Non-invasive fetal ECG analysis. *Physiological Measurement*, 35(8):1521, 2014.
- [77] F. V. Coakley, H. Hricak, R. A. Filly, A. J. Barkovich, and M. R. Harrison. Complex fetal disorders: effect of MR imaging on management—preliminary clinical experience. *Radiology*, 213(3):691–696, 1999.
- [78] W. R. Cohen, S. Ommani, S. Hassan, F. G. Mirza, M. Solomon, R. Brown, B. S. Schifrin, J. M. Himsworth, and B. R. Hayes-Hill. Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. *Acta Obstetrica et Gynecologica Scandinavica*, 91(11):1306–1313, 2012.
- [79] W. Count. A quantitative analysis of growth in certain human skull dimensions. *Human Biology*, 14(2):143–165, 1942.
- [80] M. Covas, E. Alda, S. Ventura, S. Braunstein, G. Serralunga, and L. Yañez. Variación del peso durante el primer mes de vida en recién nacidos de término sanos con lactancia materna exclusiva. *Archivos Argentinos de Pediatría*, 104(5):399–405, 2006.
- [81] 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.
- [82] S. Cullen, G. K. Sharland, L. D. Allan, and I. D. Sullivan. Potential impact of population screening for prenatal diagnosis of congenital heart disease. *Archives of Disease in Childhood*, 67(7 Spec No):775–778, 1992.
- [83] G. Dawes, C. Houghton, and C. Redman. Baseline in human fetal heart-

- rate records. *BJOG: An International Journal of Obstetrics and Gynaecology*, 89(4):270–275, 1982.
- [84] 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.
- [85] 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.
- [86] J. V. de Bernabé, T. Soriano, R. Albaladejo, M. Juarranz, M. E. Calle, and V. Martínez, David an Domínguez-Rojas. Risk factors for low birth weight: a review. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 116(1):3–15, 2004.
- [87] L. de Bernis, M. V. Kinney, W. Stones, P. ten Hoop-Bender, D. Vivio, S. H. Leisher, Z. A. Bhutta, M. Gülmezoglu, M. Mathai, J. M. Belizán, and L. Franco. Stillbirths: ending preventable deaths by 2030. *The Lancet*, 387(10019):703–716, 2016.
- [88] J. de Laveaucoupet, F. Audibert, F. Guis, C. Rambaud, B. Suarez, C. Boithias-Guérot, and D. Musset. Fetal magnetic resonance imaging (MRI) of ischemic brain injury. *Prenatal Diagnosis*, 21(9):729–736, 2001.
- [89] 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.
- [90] J. De Wilde, A. Rivers, and D. Price. A review of the current use of magnetic resonance imaging in pregnancy and safety implications for the fetus. *Progress in Biophysics and Molecular Biology*, 87(2-3):335–353, 2005.

- [91] 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.
- [92] 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.
- [93] D. Devane, J. G. Lalor, S. Daly, W. McGuire, and V. Smith. Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing. *Cochrane Database Systematic Revision*, 2, 2012.
- [94] 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.
- [95] 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.
- [96] P. Diggle, P. J. Diggle, P. Heagerty, K.-Y. Liang, P. J. Heagerty, and S. Zeger. *Analysis of longitudinal data*. Oxford University Press, Oxford, UK, 2002.
- [97] D. DiTomasso and A. L. Paiva. Neonatal weight matters: An examination of weight changes in full-term breastfeeding newborns during the first 2 weeks of life. *Journal of Human Lactation*, 34(1):86–92, 2018.
- [98] D. DiTomasso, M. Roberts, and B. P. Cotton. Postpartum mothers’ experiences

- with newborn weight checks in the home. *The Journal of Perinatal and Neonatal Nursing*, 32(4):333–340, 2018.
- [99] J. Dobbe, S. Lunshof, K. Boer, H. Wolf, and C. Grimbergen. The technique and algorithms for computerized analysis of long-term fetal heart rate recordings. *Prenatal and Neonatal Medicine*, 6(5):280–289, 2001.
- [100] A. Dražančić. Antenatal care in developing countries. what should be done? *Journal of Perinatal Medicine*, 29(3):188–198, 2001.
- [101] L. Duley. The global impact of pre-eclampsia and eclampsia. In *Seminars in Perinatology*, volume 33, pages 130–137. Elsevier, 2009.
- [102] R. D. Eden. Postdate pregnancy: antenatal assessment of fetal well-being. *Clinical Obstetrics and Gynecology*, 32(2):235–244, 1989.
- [103] R. D. Eden, F. H. Boehm, and M. Haire. Assessment and care of the fetus: Physiological, clinical, and medicolegal principles. *JAMA: The Journal of the American Medical Association*, 264(18), 1990.
- [104] S. H. Eik-Nes. Chapter 1 - physics and instrumentation. In J. W. Wladimiroff and S. H. Eik-Nes, editors, *Ultrasound in Obstetrics and Gynecology*, pages 1 – 20. Elsevier, Edinburgh, 2009.
- [105] T. Y. Euliano, S. Darmanjian, M. T. Nguyen, J. D. Busowski, N. Euliano, and A. R. Gregg. Monitoring fetal heart rate during labor: a comparison of three methods. *Journal of Pregnancy*, 2017, 2017.
- [106] D. H. Evans. *Doppler ultrasound: Physics, instrumentation, and clinical applications*. John Wiley & Sons, Hoboken, NJ, 1989.

- [107] 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.
- [108] Fandroide. FFT spectral analyzer, 2016. Retrieved from <https://github.com/gilpanal/FFTSpectralAnalyzer> on Mar 15 2018.
- [109] A. Fanelli, G. Magenes, M. Campanile, and M. G. Signorini. Quantitative assessment of fetal well-being through ctg recordings: a new parameter based on phase-rectified signal average. *IEEE Journal of Biomedical and Health Informatics*, 17(5):959–966, 2013.
- [110] 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.
- [111] M. Ferrario, M. G. Signorini, and G. Magenes. Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. *Medical and Biological Engineering and Computing*, 47(9):911–919, 2009.
- [112] 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.
- [113] 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.
- [114] D. Fitzgerald and J. Drumm. Non-invasive measurement of human fetal circulation using ultrasound: a new method. *BMJ*, 2(6100):1450–1451, 1977.

- [115] V. Flaherman, E. W. Schaefer, M. W. Kuzniewicz, S. X. Li, E. M. Walsh, and I. M. Paul. Health care utilization in the first month after birth and its relationship to newborn weight loss and method of feeding. *Academic Pediatrics*, 18(6):677–684, 2018.
- [116] V. J. Flaherman, E. W. Schaefer, M. W. Kuzniewicz, S. X. Li, E. M. Walsh, and I. M. Paul. Early weight loss nomograms for exclusively breastfed newborns. *Pediatrics*, 135(1):e16–e23, 2015.
- [117] 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.
- [118] K. B. Fortner. *The Johns Hopkins manual of gynecology and obstetrics*. Lippincott Williams & Wilkins, Philadelphia, PA, 2007.
- [119] M. C. Frates, A. J. Kumar, C. B. Benson, V. L. Ward, and C. M. Tempny. Fetal anomalies: comparison of MR imaging and us for diagnosis. *Radiology*, 232(2):398–404, 2004.
- [120] 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.
- [121] R. K. Freeman, T. J. Garite, M. P. Nageotte, and L. A. Miller. *Fetal heart rate monitoring*. Lippincott Williams & Wilkins, Philadelphia, PA, 2012.
- [122] 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.

- [123] K. B. Gan, E. Zahedi, and M. A. M. Ali. Transabdominal fetal heart rate detection using NIR photoplethysmography: instrumentation and clinical results. *IEEE transactions on Biomedical Engineering*, 56(8):2075–2082, 2009.
- [124] 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.
- [125] J. Gardosi, V. Madurasinghe, M. Williams, A. Malik, and A. Francis. Maternal and fetal risk factors for stillbirth: population based study. *BMJ*, 346:f108, 2013.
- [126] L. Geerts, A. Theron, D. Grove, G. Theron, and H. Odendaal. A community-based obstetric ultrasound service. *International Journal of Gynecology and Obstetrics*, 84(1):23–31, 2004.
- [127] L. T. Geerts, E. J. Brand, and G. B. Theron. Routine obstetric ultrasound examinations in South Africa: cost and effect on perinatal outcome—a prospective randomised controlled trial. *BJOG: An International Journal of Obstetrics and Gynaecology*, 103(6):501–507, 1996.
- [128] M. A. Gelbart, J. Snoek, and R. P. Adams. Bayesian optimization with unknown constraints. *arXiv preprint arXiv:1403.5607*, 2014.
- [129] 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.

- [130] O. Glenn and A. Barkovich. Magnetic resonance imaging of the fetal brain and spine: an increasingly important tool in prenatal diagnosis, part 1. *American Journal of Neuroradiology*, 27(8):1604–1611, 2006.
- [131] O. A. Glenn, R. B. Goldstein, K. C. Li, S. J. Young, M. E. Norton, R. F. Busse, J. D. Goldberg, and A. J. Barkovich. Fetal magnetic resonance imaging in the evaluation of fetuses referred for sonographically suspected abnormalities of the corpus callosum. *Journal of Ultrasound in Medicine*, 24(6):791–804, 2005.
- [132] 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.
- [133] 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 Obstetrica et Gynecologica Scandinavica*, 86(11):1303–1309, 2007.
- [134] R. L. Goldenberg, E. M. McClure, and S. Saleem. Improving pregnancy outcomes in low-and middle-income countries. *Reproductive Health*, 15(1):88, 2018.
- [135] R. L. Goldenberg, E. M. McClure, S. Saleem, and U. M. Reddy. Infection-related stillbirths. *The Lancet*, 375(9724):1482–1490, 2010.
- [136] R. L. Goldenberg, S. Saleem, O. Pasha, M. S. Harrison, and E. M. McClure. Reducing stillbirths in low-income countries. *Acta Obstetrica et Gynecologica Scandinavica*, 95(2):135–143, 2016.
- [137] R. A. Grant, G. G. Heuer, G. M. Carrión, N. S. Adzick, E. S. Schwartz, S. C. Stein, P. B. Storm, and L. N. Sutton. Morphometric analysis of posterior fossa after in utero myelomeningocele repair. *Journal of Neurosurgery: Pediatrics*, 7(4):362–368, 2011.

- [138] V. Grech, C. Savona-Ventura, and P. Vassallo-Agius. Unexplained differences in sex ratios at birth in Europe and North America. *BMJ*, 324(7344):1010–1011, 2002.
- [139] A. Griffiths, A. Pinto, and L. Margarit. A survey of methods used to measure symphysis fundal height. *J Obstet Gynaecol*, 28(7):692–694, oct 2008.
- [140] R. M. Grivell, Z. Alfirevic, G. M. Gyte, and D. Devane. Antenatal cardiotocography for fetal assessment. *Cochrane Database of Systematic Reviews*, (9), 2015.
- [141] 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.
- [142] A. B. Hameed and M. S. Sklansky. Pregnancy: maternal and fetal heart disease. *Current Problems in Cardiology*, 32(8):419–494, 2007.
- [143] P. Hamelmann, R. Vullings, A. F. Kolen, J. W. Bergmans, J. O. van Laar, P. Tortoli, and M. Mischi. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2019.
- [144] 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.
- [145] 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.
- [146] R. C. Hauspie and L. Molinari. Parametric models for postnatal growth. In R. C. Hauspie, N. Cameron, and L. Molinari, editors, *Methods in Human Growth*

- Research*, chapter 8, pages 209–212. Cambridge University Press, Cambridge, UK, 2004.
- [147] H. He, Y. Bai, E. A. Garcia, and S. Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1322–1328. IEEE, 2008.
- [148] T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283, 1988.
- [149] L. Hill, I. Sollers, and J. Thayer. Evaluation of a simple estimation method for the derivation of cardiac output from arterial blood pressure and heart rate. *Biomedical Science Instrumentation*, 48:165–170, 2011.
- [150] M. Hladunewich, S. A. Karumanchi, and R. Lafayette. Pathophysiology of the clinical manifestations of preeclampsia. *Clinical Journal of the American Society of Nephrology*, 2(3):543–549, 2007.
- [151] 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.
- [152] E. Hon. The instrumentation of fetal heart rate and fetal electrocardiography. I. A fetal heart monitor. *Connecticut Medicine*, 24:289, 1960.
- [153] E. Hon and O. Hess. The clinical value of fetal electrocardiography. *American Journal of Obstetrics and Gynecology*, 79:1012–1023, 1960.
- [154] E. Hon and S. Lee. Averaging techniques in fetal electrocardiography. *Medical Electronics and Biological Engineering*, 2(1):71–76, 1964.

- [155] 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.
- [156] D. Hoyer, S. Nowack, S. Bauer, F. Tetschke, A. Rudolph, U. Wallwitz, F. Jaenicke, E. Heinicke, T. Götz, R. Huonker, and O. Witte. Fetal development of complex autonomic control evaluated from multiscale heart rate patterns. *American Journal of Physiology-Heart and Circulatory Physiology*, 304(5):R383–R392, 2012.
- [157] 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.
- [158] N. E. Huang and S. S. P. Shen. *Hilbert-Huang transform and its applications*, volume Interdisciplinary Mathematical Sciences: Volume 16. World Scientific, Singapore, 2 edition, 2014.
- [159] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998.
- [160] X. Huang, A. Acero, H.-W. Hon, and R. Reddy. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, volume 95. Prentice Hall PTR Upper Saddle River, 2001.

- [161] A. M. Hubbard. Magnetic resonance imaging of fetal thoracic abnormalities. *Topics in Magnetic Resonance Imaging*, 12(1):18–24, 2001.
- [162] A. M. Hubbard. Ultrafast fetal MRI and prenatal diagnosis. *Seminars in Pediatric Surgery*, 12(3):143–153, 2003.
- [163] E. Huhn, S. Lobmaier, T. Fischer, R. Schneider, A. Bauer, K. Schneider, and G. Schmidt. New computerized fetal heart rate analysis for surveillance of intrauterine growth restriction. *Prenatal Diagnosis*, 31(5):509–514, 2011.
- [164] K. J. Hunt and K. L. Schuller. The increasing prevalence of diabetes in pregnancy. *Obstetrics and gynecology clinics of North America*, 34(2):173–199, 2007.
- [165] G. Hypertension. Preeclampsia. acog practice bulletin no. 202. american college of obstetricians and gynecologists. *Obstetrics and Gynecology*, 133(1):e1–e25, 2019.
- [166] A. A. Ibarra-Polo, E. F. Guiloff, and C. Gomez-Rogers. Fetal heart rate throughout pregnancy. *American Journal of Obstetrics and Gynecology*, 113(6):814–818, 1972.
- [167] C. Ionescu. The benefits of 3D-4D fetal echocardiography. *Maedica*, 5(1):45, 2010.
- [168] P. A. Janssen, P. Thiessen, M. C. Klein, M. F. Whitfield, Y. C. MacNab, and S. C. Cullis-Kuhl. Standards for the measurement of birth weight, length and head circumference at term in neonates of European, Chinese and South Asian ancestry. *Open Medicine*, 1(2):e74, 2007.
- [169] 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 (ICACISIS)*, pages 29–36. IEEE, 2015.

- [170] 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.
- [171] R. M. Jenss and N. Bayley. A mathematical method for studying the growth of a child. *Human Biology*, 9(4):556, 1937.
- [172] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern Recognition*, 44(9):2231–2240, 2011.
- [173] J. Jezewski, T. Kupka, and K. Horoba. Extraction of fetal heart-rate signal as the time event series from evenly sampled data acquired using Doppler ultrasound technique. *IEEE Transactions on Biomedical Engineering*, 55(2):805–810, 2008.
- [174] 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.
- [175] 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.
- [176] 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, 2017.

- [177] 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.
- [178] G. Kac, R. H. Mendes, D. R. Farias, I. Eshriqui, F. Rebelo, C. Benaim, A. A. F. Vilela, N. S. Lima, W. A. Peres, and G. F. Salles. Hepatic, renal and inflammatory biomarkers are positively associated with blood pressure changes in healthy pregnant women: a prospective cohort. *Medicine*, 94(20), 2015.
- [179] 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.
- [180] 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.
- [181] S. Karl, C. S. L. W. Suen, H. W. Unger, M. Ome-Kaius, G. Mola, L. White, R. A. Wangnapi, S. J. Rogerson, and I. Mueller. Preterm or not—an evaluation of estimates of gestational age in a cohort of women from rural Papua New Guinea. *PloS One*, 10(5):e0124286, 2015.
- [182] K. S. Khan, D. Wojdyla, L. Say, A. M. Gülmezoglu, and P. F. Van Look. Who analysis of causes of maternal death: a systematic review. *The Lancet*, 367(9516):1066–1074, 2006.
- [183] 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.
- [184] A. Khandoker, C. Karmakar, Y. Kimura, M. Endo, S. Oshio, and M. Palaniswami. Tone entropy analysis of foetal heart rate variability. *Entropy*, 17(3):1042–1053, 2015.
- [185] A. H. Khandoker, Y. Kimura, T. Ito, N. Sato, K. Okamura, and M. Palaniswami. Antepartum non-invasive evaluation of opening and closing timings of the cardiac valves in fetal cardiac cycle. *Medical and Biological Engineering and Computing*, 47(10):1075–1082, 2009.
- [186] I. Kim and E. Pollitt. Differences in the pattern of weight growth of nutritionally at-risk and well-nourished infants. *The American Journal of Clinical Nutrition*, 46(1):31–35, 1987.
- [187] A. Kluttig, D. Tiller, B. Schumann, O. Kuss, K. Greiser, K. Werdan, and J. Haerting. Blood pressure measurement: differences between arm side, sitting and supine position and between consecutive measurements. *Das Gesundheitswesen*, 72(08/09):V215, 2010.
- [188] R. Koenker and B. J. Park. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71(1-2):265–283, 1996.
- [189] 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.
- [190] A. P. Kourtis, J. S. Read, and D. J. Jamieson. Pregnancy and infection. *New England Journal of Medicine*, 370(23):2211–2218, 2014.

- [191] 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.
- [192] D. Kumar, P. Carvalho, M. Antunes, R. Paiva, and J. Henriques. Noise detection during heart sound recording using periodicity signatures. *Physiological Measurement*, 32(5):599, 2011.
- [193] L. N. LaGrone, V. Sadasivam, A. L. Kushner, and R. S. Groen. A review of training opportunities for ultrasonography in low and middle income countries. *Tropical Medicine and International Health*, 17(7):808–819, 2012.
- [194] 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.
- [195] I. Lakhno. The impact of preeclampsia on fetal ecg morphology and heart rate variability. *Archives of Perinatal Medicine*, 20(1):7–10, 2014.
- [196] I. Lakhno. Autonomic imbalance captures maternal and fetal circulatory response to pre-eclampsia. *Clinical Hypertension*, 23(1):5, 2017.
- [197] D. Lane. Online statistics education: A multimedia course of study. In *EdMedia: World Conference on Educational Media and Technology*, pages 1317–1320. Association for the Advancement of Computing in Education (AACE), 2003.
- [198] S. Lange, P. Van Leeuwen, D. Geue, W. Hatzmann, and D. Grönemeyer. Influence of gestational age, heart rate, gender and time of day on fetal heart rate variability. *Medical and Biological Engineering and Computing*, 43(4):481–486, 2005.

- [199] S. Larks and G. Larks. Components of the fetal electrocardiogram and intrauterine electrical axis: Quantitative data. *Neonatology*, 10(3-4):140–152, 1966.
- [200] S. D. Larks and L. D. Longo. Electrocardiographic studies of the fetal heart during delivery. *Obstetrics and Gynecology*, 19(6):740–747, 1962.
- [201] N. H. Lauersen, H. M. Hochberg, and M. E. George. Evaluation of the accuracy of a new ultrasonic fetal heart rate monitor. *American Journal of Obstetrics and Gynecology*, 125(8):1125–1135, 1976.
- [202] 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.
- [203] 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.
- [204] J. E. Lawn, A. C. Lee, M. Kinney, L. Sibley, W. A. Carlo, V. K. Paul, R. Patinson, 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.
- [205] 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.

- [206] 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.
- [207] 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.
- [208] D. Levine, P. D. Barnes, and R. R. Edelman. Obstetric MR imaging. *Radiology*, 211(3):609–617, 1999.
- [209] Q. Li and G. Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491–1503, 2012.
- [210] Q. Li and G. D. Clifford. Signal quality and data fusion for false alarm reduction in the intensive care unit. *Journal of Electrocardiology*, 45(6):596–603, 2012.
- [211] T. Li, T. Qiu, and H. Tang. Optimum heart sound signal selection based on the cyclostationary property. *Computers in Biology and Medicine*, 43(6):607–612, 2013.
- [212] T. Li, H. Tang, T. Qiu, and Y. Park. Best subsequence selection of heart sound recording based on degree of sound periodicity. *Electronics Letters*, 47(15):841–843, 2011.
- [213] R. Liston, D. Sawchuck, and D. Young. Fetal health surveillance: antepartum and intrapartum consensus guideline. *Journal of Obstetrics and Gynaecology*

- Canada: *JOGC= Journal d'Obstétrique et Gynécologie du Canada: JOGC*, 29(9 Suppl 4):S3–56, 2007.
- [214] C. Liu, D. Springer, Q. Li, B. Moody, R. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson, Z. Syed, S. Schmidt, c. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, S. M. R., R. Sameni, R. Mark, and G. D. Clifford. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37(12):2181, 2016.
- [215] 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.
- [216] F. Lunghi, G. Magenes, L. Pedrinazzi, and M. G. Signorini. Detection of fetal distress though a support vector machine based on fetal heart rate parameters. In *Computers in Cardiology, 2005*, pages 247–250. IEEE, 2005.
- [217] P. Macdonald, S. Ross, L. Grant, and D. Young. Neonatal weight loss in breast and formula fed infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 88(6):F472–F476, 2003.
- [218] S. N. MacGregor and R. Sabbagha. Assessment of gestational age by ultrasound. *The Global Library of Women's Medicine*, 2008.
- [219] E. F. Magann, S. P. Chauhan, P. S. Barrilleaux, N. S. Whitworth, and J. N. Martin Jr. Amniotic fluid index and single deepest pocket: weak indicators of abnormal amniotic volumes. *Obstetrics and Gynecology*, 96(5):737–740, 2000.
- [220] 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.

- [221] 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.
- [222] 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.
- [223] A. Maitra and N. Kuntagod. A novel mobile application to assist maternal health workers in rural India. In *2013 5th International Workshop on Software Engineering in Health Care (SEHC)*, pages 75–78. IEEE, 2013.
- [224] 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.
- [225] P. Malcus. Antenatal fetal surveillance. *Current Opinion in Obstetrics and Gynecology*, 16(2):123–128, 2004.
- [226] 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.
- [227] L. Mangesi, G. J. Hofmeyr, V. Smith, and R. M. Smyth. Fetal movement counting for assessment of fetal wellbeing. *Cochrane Database of Systematic Reviews*, (10), 2015.
- [228] R. Mantel, H. Van Geijn, F. Caron, J. Swartjes, E. Van Woerden, and H. Jongsma. Computer analysis of antepartum fetal heart rate: 1. Baseline de-

- termination. *International Journal of Bio-Medical Computing*, 25(4):261–272, 1990.
- [229] G. Mari and F. Hanif. Fetal Doppler: umbilical artery, middle cerebral artery, and venous system. *Seminars in Perinatology*, 32(4):253–257, 2008.
- [230] E. N. Marieb and K. Hoehn. *Human Anatomy & Physiology*. Pearson Education, Boston, MA, USA, 2007.
- [231] 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.
- [232] 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.
- [233] 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.
- [234] F. Marzbanrad. *Modeling fetal cardiac valve intervals and fetal-maternal interactions*. PhD thesis, The University of Melbourne, Melbourne, Australia, 2015.
- [235] F. Marzbanrad, A. Khandoker, Y. Kimura, M. Palaniswami, and G. D. Clifford. Estimating fetal gestational age using cardiac valve intervals. In *Computing in Cardiology Conference (CinC)*. IEEE, 2016.

- [236] F. Marzbanrad, A. H. Khandoker, B. D. Hambly, E. Ng, M. Tamayo, Y. Lu, S. Matthews, C. Karmakar, M. Palaniswami, H. F. Jelinek, and C. McLachlan. Methodological comparisons of heart rate variability analysis in patients with type 2 diabetes and angiotensin converting enzyme polymorphism. *IEEE Journal of Biomedical and Health Informatics*, 20(1):55–63, 2016.
- [237] 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.
- [238] F. Marzbanrad, Y. Kimura, M. Endo, M. Palaniswami, and A. H. Khandoker. Classification of Doppler ultrasound signal quality for the application of fetal valve motion identification. *Computing in Cardiology Conference (CinC), 2015*, 42:365–368, 2015.
- [239] 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, 2016.
- [240] 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.
- [241] F. Marzbanrad, L. Stroux, and G. D. Clifford. Cardiotocography and beyond: a review of one-dimensional Doppler ultrasound application in fetal monitoring. *Physiological Measurement*, 2018.
- [242] S. Matsuoka, K. Takeuchi, Y. Yamanaka, Y. Kaji, K. Sugimura, and T. Maruo. Comparison of magnetic resonance imaging and ultrasonography in the prenatal

- diagnosis of congenital thoracic abnormalities. *Fetal Diagnosis and Therapy*, 18(6):447–453, 2003.
- [243] E. M. McClure, R. O. Nathan, S. Saleem, F. Esamai, A. Garces, E. Chomba, A. Tshefu, D. Swanson, H. Mabeya, L. Figuero, and W. Mirza. First look: a cluster-randomized trial of ultrasound to improve pregnancy outcomes in low income country settings. *BMC Pregnancy and Childbirth*, 14(1):1, 2014.
- [244] 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.
- [245] P. J. Meis, M. Klebanoff, E. Thom, M. P. Dombrowski, B. Sibai, A. H. Moawad, C. Y. Spong, J. C. Hauth, M. Miodovnik, M. W. Varner, and K. Leveno. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *New England Journal of Medicine*, 348(24):2379–2385, 2003.
- [246] E. Merz. *Ultrasound in obstetrics and gynecology*. Georg Thieme Verlag, Stuttgart, Germany, 2004.
- [247] 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.
- [248] J. R. Miller, V. J. Flaherman, E. W. Schaefer, M. W. Kuzniewicz, S. X. Li, E. M. Walsh, and I. M. Paul. Early weight loss nomograms for formula fed newborns. *Hospital Pediatrics*, 5(5):263–268, 2015.
- [249] A. Mohr Sasson, A. Tsur, A. Kalter, A. Weissmann Brenner, L. Gindes, and B. Weisz. Reduced fetal movement: factors affecting maternal perception. *The Journal of Maternal-Fetal and Neonatal Medicine*, 29(8):1318–1321, 2016.

- [250] 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.
- [251] T. R. Moore and K. Piacquadio. A prospective evaluation of fetal movement screening to reduce the incidence of antepartum fetal death. *American Journal of Obstetrics and Gynecology*, 160(5):1075–1080, 1989.
- [252] 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.
- [253] R. Morris, R. Say, S. Robson, J. Kleijnen, and K. Khan. Systematic review and meta-analysis of middle cerebral artery doppler to predict perinatal wellbeing. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 165(2):141–155, 2012.
- [254] MSPAS/Guatemala, INE/Guatemala, Segeplán/Guatemala, and ICF International. Encuesta nacional de salud materno infantil 2014-2015: Informe final, 2017. Retrieved from <https://dhsprogram.com/publications/publication-fr318-dhs-final-reports.cfm> on Oct 11 2019.
- [255] Y. Murata and J. Chester B Martin. Systolic time intervals of the fetal cardiac cycle. *Obstetrics and Gynecology*, 44(2):224–232, 1974.
- [256] Y. Murata and C. B. Martin Jr. Systolic time intervals of the fetal cardiac cycle. *Postgraduate Medicine*, 61(4):127–134, 1977.
- [257] R. Mustafa, S. Ahmed, A. Gupta, and R. C. Venuto. A comprehensive review of hypertension in pregnancy. *Journal of Pregnancy*, 2012, 2012.
- [258] M. K. Mwaniki, M. Atieno, J. E. Lawn, and C. R. Newton. Long-term neu-

- rodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *The Lancet*, 379(9814):445–452, 2012.
- [259] National Institute of Child Health and Human Development. What are some common complications of pregnancy?, 2017. Retrieved from <http://www.who.int/mediacentre/factsheets/fs348/en/> on March 24 2020.
- [260] M. G. Neerhof. Causes of intrauterine growth restriction. *Clinics in Perinatology*, 22(2):375–385, 1995.
- [261] J. P. Neilson. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database of Systematic Reviews*, (4), 1998.
- [262] S. Neldam. Fetal movements as an indicator of fetal wellbeing. *The Lancet*, 315(8180):1222–1224, 1980.
- [263] R. Netea, J. Lenders, P. Smits, and T. Thien. Both body and arm position significantly influence blood pressure measurement. *Journal of Human Hypertension*, 17(7):459, 2003.
- [264] L. Neufeld, J. Haas, R. Grajeda, and R. Martorell. Ultrasound measurement of fetal size in rural guatemala. *International Journal of Gynecology and Obstetrics*, 84(3):220–228, 2004.
- [265] L. M. Neufeld, J. D. Haas, R. Grajéda, and R. Martorell. Last menstrual period provides the best estimate of gestation length for women in rural Guatemala. *Paediatric and Perinatal Epidemiology*, 20(4):290–298, 2006.
- [266] K. Nicolaides, G. Rizzo, K. Hecher, and R. Ximenes. The Fetal Medicine Foundation, 2002.
- [267] I. J. Nijhuis, J. ten Hof, E. J. Mulder, J. G. Nijhuis, H. Narayan, D. J. Taylor, and G. H. Visser. Fetal heart rate in relation to its variation in normal and

- growth retarded fetuses. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 89(1):27–33, 2000.
- [268] H. Norén, I. Amer-Wählin, H. Hagberg, A. Herbst, I. Kjellmer, K. Marsál, 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.
- [269] 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.
- [270] NOTEBOOKCHECK. Comparison of smartphone and tablet processor performance, 2018. Retrieved from <https://www.notebookcheck.net/Smartphone-Processors-Benchmark-List.149513.0.html> on Mar 16 2018.
- [271] E. L. Nyboe. An algorithm based on the Dawes/Redman criteria for automated fetal heart rate analysis. Master’s thesis, Chalmerss University of Technology, Göteborg, Sweden, 2011.
- [272] Office on Women’s Health. Pregnancy: pregnancy complications, 2010. Retrieved from <http://www.womenshealth.gov/pregnancy/you-are-pregnant/pregnancy-complications.html> on Jan 11 2020.
- [273] A. Oliver and C. Overton. Diagnosis and management of miscarriage. *The Practitioner*, 258(1771):25–8, 2014.
- [274] OpenStaxCollege. *Anatomy and Physiology II*. Rice University, Houston, TX, USA, 2015.

- [275] 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.
- [276] T. Pais and S. Gutiérrez. Crecimiento de los niños amamantados en el primer mes de vida. *Revista Médica del Uruguay*, 19(3):201–207, 2003.
- [277] E. K. Pallotto and H. W. Kilbride. Perinatal outcome and later implications of intrauterine growth restriction. *Clinical Obstetrics and Gynecology*, 49(2):257–269, 2006.
- [278] 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.
- [279] 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.
- [280] J. T. Parer, E. J. Quilligan, F. H. Boehm, R. Depp, L. D. Devoe, M. Y. Divon, K. R. Greene, C. J. Harvey, J. C. Hauth, and J. F. Huddleston. Electronic fetal heart rate monitoring: research guidelines for interpretation. *American Journal of Obstetrics and Gynecology*, 177(6):1385–1390, 1997.
- [281] 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.

- [282] I. M. Paul, E. W. Schaefer, J. R. Miller, M. W. Kuzniewicz, S. X. Li, E. M. Walsh, and V. J. Flaherman. Weight change nomograms for the first month after birth. *Pediatrics*, 138(6):e20162625, 2016.
- [283] V. K. Paul and M. Singh. Regionalized perinatal care in developing countries. *Seminars in Neonatology*, 9(2):117–124, 2004.
- [284] J. M. Peerson, M. J. Heinig, L. A. Nommsen, B. Lönnerdal, and K. G. Dewey. Use of growth models to describe patterns of length, weight, and head circumference among breast-fed and formula-fed infants: the Darling study. *Human Biology*, 65(4):611–626, 1993.
- [285] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.
- [286] I. Peter, K. Yakovenko, and G. Livshits. Relationship between parameters of early growth in israeli infants. *Homo*, 53(2):146–156, 2002.
- [287] J. R. Peter, J. J. Ho, J. Valliapan, and S. Sivasangari. Symphysial fundal height (sfh) measurement in pregnancy for detecting abnormal fetal growth. *Cochrane Database of Systematic Reviews*, (9), 2015.
- [288] C. Peters, E. ten Broeke, P. Andriessen, B. Vermeulen, R. Berendsen, P. Wijn, and S. Oei. Beat-to-beat detection of fetal heart rate: Doppler ultrasound cardiotocography compared to direct ECG cardiotocography in time and frequency domain. *Physiological measurement*, 25(2):585, 2004.
- [289] 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.

- [290] J. Phelan, M. Ahn, C. V. Smith, S. Rutherford, and E. Anderson. Amniotic fluid index measurements during pregnancy. *The Journal of Reproductive Medicine*, 32(8):601–604, 1987.
- [291] 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.
- [292] 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.
- [293] J. Poutamo, R. Vanninen, K. Partanen, and P. Kirkinen. Diagnosing fetal urinary tract abnormalities: benefits of MRI compared to ultrasonography. *Acta obstetricia et gynecologica Scandinavica*, 79(1):65–71, 2000.
- [294] T. M. Quinn, A. M. Hubbard, and N. S. Adzick. Prenatal magnetic resonance imaging enhances fetal diagnosis. *Journal of Pediatric Surgery*, 33(4):553–558, 1998.
- [295] U. Ramakrishnan. Nutrition and low birth weight: from research to practice. *The American Journal of Clinical Nutrition*, 79(1):17–21, 2004.
- [296] S. Raymond and C. Whitfield. 10 systolic time intervals of the fetal cardiac cycle. *Baillière’s Clinical Obstetrics and Gynaecology*, 1(1):185–201, 1987.
- [297] 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.

- [298] 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.
- [299] J. Reinhard, B. R. Hayes-Gill, S. Schiermeier, H. Hatzmann, T. M. Heinrich, and F. Louwen. Intrapartum heart rate ambiguity: a comparison of cardiotocogram and abdominal fetal electrocardiogram with maternal electrocardiogram. *Gynecologic and Obstetric Investigation*, 75(2):101–108, 2013.
- [300] J. Reinhard, B. R. Hayes-Gill, Q. Yi, H. Hatzmann, and S. Schiermeier. Comparison of non-invasive fetal electrocardiogram to Doppler cardiotocogram during the 1st stage of labor. *Journal of Perinatal Medicine*, 38(2):179–185, 2010.
- [301] M. Resta, P. Greco, V. D’Addario, C. Florio, N. Dardes, G. Caruso, P. Spagnolo, R. Clemente, A. Vimercati, and L. Selvaggi. Magnetic resonance imaging in pregnancy: study of fetal cerebral malformations. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 4(1):7–20, 1994.
- [302] I. Rezek and S. J. Roberts. Envelope extraction via complex homomorphic filtering. *Technical Report TR-98-9 Technical report*, 1998.
- [303] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.
- [304] M. J. Rijken, A. M. De Livera, S. J. Lee, M. E. Boel, S. Rungwilailaekhiri, J. Wiladphaingern, M. K. Paw, M. Pimanpanarak, S. Pukrittayakamee, J. A. Simpson, and F. Nosten. Quantifying low birth weight, preterm birth and small-for-gestational-age effects of malaria in pregnancy: a population cohort study. *PLoS One*, 9(7):e100247, 2014.

- [305] D. Riknagel, B. Dinesen, H. Zimmermann, R. Farlie, S. Schmidt, E. Toft, and J. J. Struijk. Digital auscultation of the uterine artery: a measure of uteroplacental perfusion. *Physiological Measurement*, 37(7):1163, 2016.
- [306] O. Rioul and M. Vetterli. Wavelets and signal processing. *Signal Processing Magazine, IEEE*, 8(4):14–38, 1991.
- [307] B. F. Robinson. Relation of heart rate and systolic blood pressure to the onset of pain in angina pectoris. *Circulation*, 35(6):1073–1083, 1967.
- [308] D. Roj, J. Wróbel, K. Horoba, T. Przybyła, and T. Kupka. Improving the periodicity measurement in fetal heart activity signal. *Journal of Medical Informatics and Technologies*, 16, 2010.
- [309] S. Romagnoli, A. Sbröllini, L. Burattini, I. Marcantoni, M. Morettini, and L. Burattini. Digital cardiocography: What is the optimal sampling frequency? *Biomedical Signal Processing and Control*, 51:210–215, 2019.
- [310] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2):117–134, 1993.
- [311] Royal College of Obstetricians and Gynaecologists. Small-for-Gestational-Age Fetus, Investigation and Management (Green-top Guideline No. 31), 2002.
- [312] S. E. Rutherford, J. P. Phelan, C. V. Smith, and N. Jacobs. The four-quadrant assessment of amniotic fluid volume: an adjunct to antepartum fetal heart rate testing. *Obstetrics and Gynecology*, 70(3):353–356, 1987.
- [313] Y. Saeys, I. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

- [314] M. Saguintaah, A. Couture, C. Veyrac, C. Baud, and M.-P. Quere. MRI of the fetal gastrointestinal tract. *Pediatric Radiology*, 32(6):395–404, 2002.
- [315] S. P. Salas, G. Marshall, B. L. Gutiérrez, and P. Rosso. Time course of maternal plasma volume and hormonal changes in women with preeclampsia or fetal growth restriction. *Hypertension*, 47(2):203–208, 2006.
- [316] R. Sameni, G. Clifford, J. Ward, J. Robertson, C. Pettigrew, and A. Wolfberg. 664: Accuracy of fetal heart rate acquired from sensors on the maternal abdomen compared to a fetal scalp electrode. *American Journal of Obstetrics and Gynecology*, 201(6):S241, 2009.
- [317] 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.
- [318] H. F. Sandmire and R. K. DeMott. Electronic fetal heart rate monitoring: research guidelines for interpretation. *American Journal of Obstetrics and Gynecology*, 179(1):276–277, 1998.
- [319] S. Santo, D. Ayres-de Campos, C. Costa-Santos, W. Schnettler, A. Ugwumadu, and L. M. Da Graça. Agreement and accuracy using the FIGO, ACOG and NICE cardiotocography interpretation guidelines. *Acta Obstetrica et Gynecologica Scandinavica*, 96(2):166–175, 2017.
- [320] Save the Children. State of the world’s newborns. Technical report, World Health Organization, Washington D.C., 2001.
- [321] 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.

- [322] 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.
- [323] 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.
- [324] J. D. Seffah and R. M. Adanu. Obstetric ultrasonography in low-income countries. *Clinical Obstetrics and Gynecology*, 52(2):250–255, 2009.
- [325] N. C. Serrano. Immunology and genetic of preeclampsia. *Journal of Immunology Research*, 13(2-4):197–201, 2006.
- [326] S. G. Shaffer, C. L. Quimiro, J. V. Anderson, and R. T. Hall. Postnatal weight changes in low birth weight infants. *Pediatrics*, 79(5):702–705, 1987.
- [327] 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.
- [328] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 & 623–656, 1948.
- [329] T. Shohoji and T. Sumiya. Individual physical growth models and biological parameters of Japanese. In P. Dasgupta and R. Hauspie, editors, *Perspectives in Human Growth, Development and Maturation*, chapter 2, pages 17–32. Springer, Dordrecht, 2001.
- [330] 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.
- [331] M. G. Signorini, A. Fanelli, and G. Magenes. Monitoring fetal heart rate during pregnancy: contributions from advanced signal processing and wearable technology. *Computational and Mathematical Methods in Medicine*, 2014, 2014.
- [332] M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini. Linear and non-linear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. *IEEE Transactions on Biomedical Engineering*, 50(3):365–374, 2003.
- [333] 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, 2020.
- [334] B. Simkhada, E. R. v. Teijlingen, M. Porter, and P. Simkhada. Factors affecting the utilization of antenatal care in developing countries: systematic review of the literature. *Journal of Advanced Nursing*, 61(3):244–260, 2008.
- [335] K. B. Simondon, F. Simondon, F. Delpeuch, and A. Cornu. Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age. *American Journal of Human Biology*, 4(3):327–335, 1992.
- [336] L. Simpson, N. J. Khati, S. P. Deshmukh, K. M. Dudiak, M. G. Harisinghani, T. L. Henrichsen, B. J. Meyer, D. A. Nyberg, L. Poder, T. D. Shipp, and C. Zelop. ACR appropriateness criteria assessment of fetal well-being. *Journal of the American College of Radiology*, 13(12):1483–1493, 2016.

- [337] A. Singh, S. Ali, A. Agarwal, and R. N. Srivastava. Correlation of shock index and modified shock index with the outcome of adult trauma patients: a prospective study of 9860 patients. *North American Journal of Medical Sciences*, 6(9):450, 2014.
- [338] 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.
- [339] 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.
- [340] J. Smith, M. Murphy, and Y. Kandasamy. The IUGR infant: a case study and associated problems with IUGR infants. *Journal of Neonatal Nursing*, 19(2):46–53, 2013.
- [341] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.
- [342] 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.
- [343] P. Sonigo, A. Elmaleh, E. Brunelle, L. Fermont, A.-L. Delezoïde, and V. Mirlesse. Prenatal MRI diagnosis of fetal cerebral tuberous sclerosis. *Pediatric Radiology*, 26(1):1–4, 1996.

- [344] P. C. Sonigo, F. F. Rypens, M. Carteret, A.-L. Delezoide, and F. O. Brunelle. MR imaging of fetal cerebral anomalies. *Pediatric Radiology*, 28(4):212–222, 1998.
- [345] 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.
- [346] D. B. Springer, T. Brennan, N. Ntusi, H. Y. Abdelrahman, L. J. Zühlke, B. M. Mayosi, L. Tarassenko, and G. D. Clifford. Automated signal quality assessment of mobile phone-recorded heart sound signals. *Journal of Medical Engineering and Technology*, 40(7-8):342–355, 2016.
- [347] 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, 2016.
- [348] P. J. Steer, M. P. Little, T. Kold-Jensen, J. Chapple, and P. Elliott. Maternal blood pressure in pregnancy, birth weight, and perinatal mortality in first births: prospective study. *BMJ*, 329(7478):1312, 2004.
- [349] U. A. Stock and J. P. Vacanti. Cardiovascular physiology during fetal development and implications for tissue engineering. *Tissue Engineering*, 7(1):1–7, 2001.
- [350] G. A. Stouffer. Useful hemodynamic formulas. In *Cardiovascular Hemodynamics for the Clinician*, chapter Appendix 1, pages 289–292. Blackwell Futura, Malden, MA, USA, 1 edition, 2008.
- [351] L. Stroux. *A perinatal monitoring system for low-resource settings*. PhD thesis, University of Oxford, 2016.

- [352] 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.
- [353] 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.
- [354] 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 & Technology*, 40(7-8):356–371, 2016.
- [355] 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 Obstetrica et Gynecologica Scandinavica*, 96(11):1322–1329, 2017.
- [356] H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926.
- [357] J. X. Sun, A. T. Reisner, and R. G. Mark. A signal abnormality index for arterial blood pressure waveforms. In *2006 Computers in Cardiology*, pages 13–16, Sept 2006.
- [358] 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.
- [359] D. M. Tax and R. P. Duin. Feature scaling in support vector data descriptions. Technical report, American Association for Artificial Intelligence, 200.
- [360] 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.
- [361] J. G. Thornton. Editorial. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 99(2):149–151, Dec 2001.
- [362] 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.
- [363] H. A. Tran. Biochemical tests in pregnancy. *Australian Prescriber*, 28:98–101, 2005.
- [364] 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.
- [365] B. Tutschek, T. Zimmermann, T. Buck, and H. Bender. Fetal tissue Doppler echocardiography: detection rates of cardiac structures and quantitative assessment of the fetal heart. *Ultrasound in Obstetrics and Gynecology*, 21(1):26–32, 2003.
- [366] UNICEF and World Health Organization. Low birthweight. country, regi-

- nal and global estimates, 2004. Retrieved from https://www.unicef.org/publications/files/low_birthweight_from_EY.pdf on Jan 23 2020.
- [367] 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.
- [368] UNICEF, World Health Organization, World Bank, and UN DESA Population Division. Low birthweight. country, regional and global estimates, 2018. Retrieved from https://data.worldbank.org/indicator/SH.DYN.NMRT?most_recent_value_desc=true on Jan 21 2020.
- [369] 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.
- [370] 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.
- [371] D. Vaiman. Genetic regulation of recurrent spontaneous abortion in humans. *Biomedical Journal*, 38(1):173–199, 2014.
- [372] 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.

- [373] 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.
- [374] 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 *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS 2018)*, Jun 2018.
- [375] 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.
- [376] C. E. Valderrama, L. Stroux, D. Springer, and G. D. Clifford. Open source code for fetal heart estimator from one-dimensional Doppler ultrasound signals. <http://doi.org/10.5281/zenodo.1487995>, 2018.
- [377] D. van der Linde, E. E. Konings, M. A. Slager, M. Witsenburg, W. A. Helbing, J. J. Takkenberg, and J. W. Roos-Hesselink. Birth prevalence of congenital heart disease worldwide. *Journal of the American College of Cardiology*, 58(21):2241–2247, 2011.
- [378] P. van Dommelen, J. P. van Wouwe, J. M. Breuning-Boers, S. van Buuren, and P. H. Verkerk. Reference chart for relative weight change to detect hypernatraemic dehydration. *Archives of Disease in Childhood*, 92(6):490–494, 2007.
- [379] P. Van Leeuwen, D. Geue, S. Lange, W. Hatzmann, and D. Grönemeyer. Changes in the frequency power spectrum of fetal heart rate in the course of pregnancy. *Prenatal Diagnosis*, 23(11):909–916, 2003.

- [380] P. Van Leeuwen, S. Lange, H. Bettermann, D. Grönemeyer, and W. Hatzmann. Fetal heart rate variability and complexity in the course of pregnancy. *Early Human Development*, 54(3):259–269, 1999.
- [381] 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.
- [382] C. L. Velayo, K. Funamoto, J. N. I. Silao, Y. Kimura, and K. Nicolaidis. Evaluation of abdominal fetal electrocardiography in early intrauterine growth restriction. *Frontiers in Physiology*, 8:437, 2017.
- [383] 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.
- [384] J. Villar and J. Belizán. The relative contribution of prematurity and fetal growth retardation to low birth weight in developing and developed societies. *American Journal of Obstetrics and Gynecology*, 143(7):793–798, 1982.
- [385] 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.
- [386] R. T. Wakai. Assessment of fetal neurodevelopment via fetal magnetocardiography. *Experimental Neurology*, 190:65–71, 2004.
- [387] 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.

- [388] 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.
- [389] 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.
- [390] 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.
- [391] T. Wheeler, K. Roberts, J. Peters, and A. Murrills. Detection of fetal movement using Doppler ultrasound. *Obstetrics and Gynecology*, 70(2):251–254, 1987.
- [392] M. Whitworth, L. Bricker, and C. Mullan. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database of Systematic Reviews*, (7), 2015.
- [393] A. L. Williamson, J. Derado, B. J. Barney, G. Saunders, I. E. Olsen, R. H. Clark, and M. L. Lawson. Longitudinal BMI growth curves for surviving preterm NICU infants based on a large us sample. *Pediatrics*, 142(3):e20174169, 2018.
- [394] A. Wolfberg. The future of fetal monitoring. *Reviews in Obstetrics and Gynecology*, 5(3/4):e132–6, 2012.
- [395] 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.

- [396] World Health Organization. Child growth standards, 2006. Retrieved from https://www.who.int/childgrowth/publications/technical_report_pub/en/ on Aug 20 2019.
- [397] World Health Organization. Global nutrition targets 2025: Low birth weight policy brief, 2014. Retrieved from https://www.who.int/nutrition/publications/globaltargets2025_policybrief_lbw/en/ on Aug 20 2019.
- [398] World Health Organization. Global status report on noncommunicable diseases 2014. Technical report, World Health Organization, 2014.
- [399] 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.
- [400] 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.
- [401] World Health Organization. Maternal mortality. Fact sheet no. 348, 2016. Retrieved from <http://www.who.int/mediacentre/factsheets/fs348/en/> on Jan 11 2020.
- [402] 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.
- [403] 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.

- [404] 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.
- [405] J. Wrobel, T. Kupka, K. Horoba, A. Matonia, D. Roj, and J. Jezewski. Recognition of fetal movements—automated detection from Doppler ultrasound signals compared to maternal perception. *Journal of Medical Imaging and Health Informatics*, 5(6):1319–1326, 2015.
- [406] L. Wu, Y. Hu, X. Liu, X. Zhang, W. Chen, S. Alan, J. A. Kellum, L. R. Waitman, and M. Liu. Feature ranking in predictive models for hospital-acquired acute kidney injury. *Scientific Reports*, 8(1):1–11, 2018.
- [407] 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.
- [408] 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.
- [409] N. A. Zaidan and M. S. Salam. Mfcc global features selection in improving speech emotion recognition rate. In *Advances in Machine Learning and Signal Processing*, pages 141–153. Springer, 2016.
- [410] C. A. Zamorano-Jiménez, J. Guzmán-Bárceñas, H. A. Baptista-González, and L. A. Fernández-Carrocerá. Pérdida de peso corporal y velocidad de crecimiento postnatal en recién nacidos menores de 1,500 gramos durante su estancia en un hospital de tercer nivel de atención. *Perinatología y Reproducción Humana*, 26(3):187–193, 2012.

- [411] G. Zanconato, R. Msolomba, L. Guarenti, and M. Franchi. Antenatal care in developing countries: the need for a tailored model. In *Seminars in Fetal and Neonatal Medicine*, volume 11, pages 15–20. Elsevier, 2006.
- [412] J. Zupan. Perinatal mortality in developing countries. *New England Journal of Medicine*, 352(20):2047–2048, 2005.