2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Toward Automated Analysis of Fetal Phonocardiograms: Comparing Heartbeat Detection from Fetal Doppler and Digital Stethoscope Signals

Yuhan Chen[1], Michael D. Wilkins[1], *Student Member, IEEE*, Jeffrey Barahona[1], Alan J. Rosenbaum[2,3], Michael Daniele[1,3], *Senior Member, IEEE* and Edgar Lobaton[1], *Member, IEEE*

Abstract— **Longitudinal fetal health monitoring is essential for high-risk pregnancies. Heart rate and heart rate variability are prime indicators of fetal health. In this work, we implemented two neural network architectures for heartbeat detection on a set of fetal phonocardiogram signals captured using fetal Doppler and a digital stethoscope. We test the efficacy of these networks using the raw signals and the hand-crafted energy from the signal. The results show a Convolutional Neural Network is the most efficient at identifying the S1 waveforms in a heartbeat, and its performance is improved when using the energy of the Doppler signals. We further discuss issues, such as low Signal-to-Noise Ratios (SNR), present in the training of a model based on the stethoscope signals. Finally, we show that we can improve the SNR, and subsequently the performance of the stethoscope, by matching the energy from the stethoscope to that of the Doppler signal.**

I. Introduction

Antepartum fetal monitoring is an important part of positive postpartum outcomes, especially for at-risk pregnancies. Fetal mortality in the United States among pregnancies reaching 20 weeks gestational age occur at a rate of approximately 6 per 1000 live births [1]. Fetal wellbeing and neurodevelopmental progress can be determined through monitoring variation in fetal heart rate due to autonomic nervous system response [2]–[8]. An at-home sensor, capable of clearly identifying audible cardiac biosignals, *e.g.*, S1 peaks of the phonocardiogram (PCG), and calculating fetal heart rate variability (FHRV) without the need for a specialist to place the sensor, would be able to warn of problems more rapidly, reducing the time between diagnosis and intervention by enabling detection of adverse events at home. This is especially important when looking for infrequent adverse events that may not occur during weekly clinic visits. Improved intervention time and quicker access to emergency obstetric care directly correlates to reduced stillbirths and neonatal deaths, which could be achieved through the adoption of an easy-to-use, at-home device [9], [10]. Maternal hypoxia can lead to low birth weight, preterm delivery, small size for gestational age, neurodevelopmental delay, fetal acidemia, and perinatal mortality [11], [12]. Causes of maternal hypoxia include lung diseases, anemia, heart disease, and sleep apnea, although snoring alone is not an indicator without other comorbidities [13]. Heart disease is on the rise worldwide, and is present in approximately half of American adults [14]. Current methods for monitoring fetal impact from maternal hypoxia are limited to infrequent clinical visits for fetal Doppler FHRV measurement, fetal cardiac decelerations compared with intrauterine contractions found via cardiotocography (CTG) antepartum or intrapartum, and other perinatal observations such as cord blood pH and neurodevelopmental progress. These methods provide little recourse for corrective action due to: (1) discovery after an insult has occurred, or (2) incomplete actionable information from lack of longitudinal data.

Most of the current automatic analysis of localization and classification of heartbeats in PCG signals focus on adult PCG signals. D. Gill et al. [15] proposed a work using homomorphic filtering to extract a smooth envelop, which yields robust heartbeat detection. Then, they built a Hidden Markov Model (HMM) to analyze the features of the detected events in order to enable unsupervised learning. A Hidden Semi-Markov Model (HSMM), extended with logistic regression, was proposed by D. B. Springer et al. [16]. This method used the heartbeat detections from electrocardiogram (ECG) signals and achieved around 95.63 % F_1 score. Zhang et al. [17] proposed an approach that used Partial Least Squares Regression (PLSR) to extract the most relevant features from scaled spectrograms, and performed classification using Support Vector Machines (SVMs). The detection of heartbeat from fetal PCG signals is more challenging due to higher frequency components and lower signal to noise ratio. M. Samieinasab and R. Sameni [18] proposed a method called Single Channel Blind Source Separation (SCBSS) consisting of Empirical Mode Decomposition (EMD) and Non-negative Matrix Factorization (NMF) to extract clean fPCG signals. They also made their dataset public for other researchers. Based on this dataset, S. Tomassini et al. [19] proposed a filter based on Wavelet transform (WT) features to clean fPCG signals.

In this paper, we compare various scenarios for heartbeat detection from fetal PCG (fPCG) signals collected by our group using fetal Doppler method and digital stethoscope. We compare the detection of S1 waveforms vs. S1-S2 waveforms for various models and inputs (including raw signals and hand-crafted energy of the signal). We discuss the issues associated with the processing of stethoscope data,

*This work was supported by National Science Foundation (NSF) under award IIS-1915599 and EEC-1160483 (ERC ASSIST), and by Easton Technologies, LLC.

[1] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, USA; Corresponding author: edgar.lobaton@ncsu.edu

[2] Department of Obstetrics and Gynecology, University of North Carolina, Chapel Hill, NC 27599, USA;

[3] Joint Department of Biomedical Engineering, North Carolina State University, University of North Carolina, Chapel Hill, NC 27599, USA;

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

and propose extensions of our work aiming to reconstruct a signal with similar SNR as the Doppler signals from the stethoscope waveforms. The rest of the article is organized as follows: Section II describes our data collection efforts and data splits used for training, validation and testing; Section III presents our methodology; Section IV discusses our results, and Section V summarizes our findings and describes our plans for future work.

II. Data Collection

A study was conducted under UNC IRB Protocol #19-1965 to provide an annotated dataset of fetal heart sounds. Abdominal acoustic recordings of 20 pregnant women were taken using a digital stethoscope (Thinklabs One, CO, USA) while the subjects were in clinic for a standard fetal Doppler non-stress test. The stethoscope acquired the raw acoustic data in the frequency range of 11 to 1000 Hz with no predefined filters applied. The audio signal generated by the fetal Doppler was recorded in tandem with the signal from the stethoscope as two simultaneous input channels of a multitrack audio recorder (H5, Zoom, NY, USA). One thousand seconds of audio was recorded for each participant. Participants also self-reported age, gestational age, pre-pregnancy height and weight, and the presence of fetal structural cardiac defects. Only singleton pregnancies were accepted, and participants with known fetal cardiac murmurs were excluded. Participants were aged from 18 to 40 years old (27.3±6.4), with a gestational age from 29 to 39 weeks (35.7±2.3), and a body mass index from 18.8 to 50.9 BMI (36.5±8.0). Segments of 30 seconds within the abdominal recordings from 5 participants were annotated by a practicing board-certified obstetrician for S1 and S2 heart sounds, and systolic and diastolic silence were determined from these annotations. An additional 2 recordings were annotated by a graduate assistant under guidance. Fig. 1 illustrate some examples.

We have two configurations for training, validation and testing. The first training set is referred to as Entire-Session (ES) Leave-Out. Five of the annotated fPCG files are used for training and validation, and 2 are set aside for testing. Of the training files, 4 were annotated by the obstetrician and the last was annotated by the graduate assistant. Of the testing files, one was annotated by an obstetrician and one was annotated by the graduate assistant. The training data was further split into training and validation, the last 10% of samples of each file is used as the validation set with the remainder as the training set. The second configuration is referred to as Fraction-of-Session (FS) Leave-Out. The last 30% of samples from each file is used for testing. The remainder is split 90% for training and 10% for validation.

III. Methodology

As described earlier, our objective is the accurate extraction of heartbeats and heart rate from the PCG recordings. We setup this problem as that of binary detection of heartbeats and consider as target detection region either the S1 annotations, or the convex hull of the corresponding S1 and S2 windows. This is done to determine for which target region we get more reliable measurements. We proceed by defining the metrics, preprocessing, and models used in this section, and analyze the results for the different variants of the problem and the methodologies in Section IV.

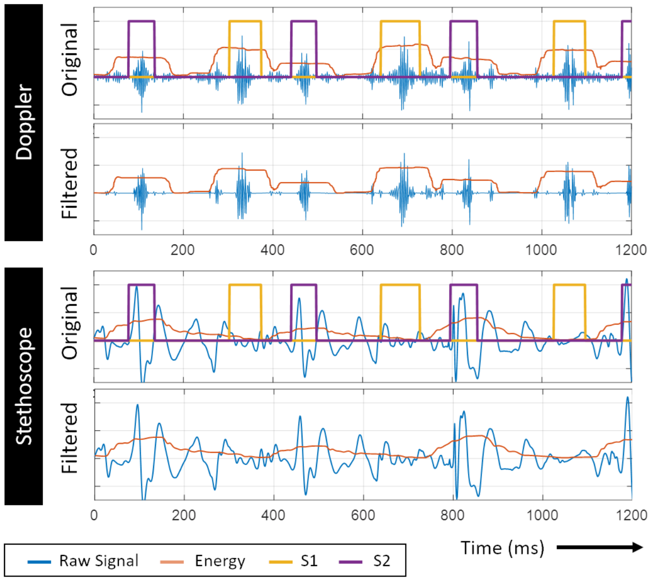

Fig. 1: Illustration of the Doppler and stethoscope audio signal after wavelet denoising and energy calculation. The ground truth labels indicating the S1 and S2 waveforms for each heartbeats are also shown. Note that denoising has a minimal effect on the stethoscope signals.

A. Metrics

In order to evaluate the performance of our heartbeat detector, we used three metrics: Precision (Prec), Recall (Rec) and the Mean Absolute Error (MAE) for the heart rate estimation. The first two metrics were computed using the number of true positives (TP), the number of true negatives (TN), the number of false positives (FP) and the number of false negatives (FN) predicted as follows:

$$\text{Prec} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

In order to obtain a heart rate for the MAE computation, we need a set of locations of the heartbeats and a window size (e.g., 4 seconds). The locations of the heartbeats are obtained by taking the middle point in a continuous segment of heartbeat detection. We take all the heartbeat locations within a window and compute the distance between consecutive heartbeats. We remove outliers using upper and lower bounds based on the training data. We used 90% of the minimum distance observed in the training data and 110% of the maximum distance observed. Finally, we consider the median of the remaining values as an estimate of heart rate over the window. The ground truth heart rate is obtained in a similar manner without outlier removal.

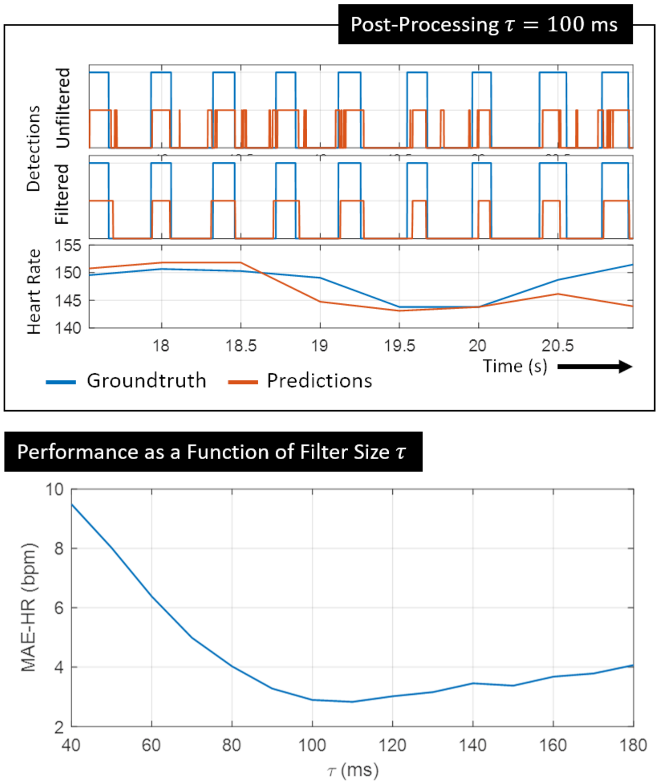

Fig. 2: Illustration of heartbeat filtering and heart rate computation. [Top] Filtering results by applying a median filter with a filter size $\tau = 100$ ms including estimated heart rate. [Bottom] The MAE for heart rate from the training data as a function of the filter size τ. A value around $\tau = 100$ ms seems to be optimal in this case yielding a training error under 3 beats per minute.

We use the median value over a window to report the estimate of heart rate since instantaneous heart rate values are more difficult to analyze. Specifically, missed detections of heartbeats can cause missing values in our data. The window size is empirically selected to be 4 seconds since this is the smallest window for which we observed that the outlier removal process does not return empty sets for the training data. Having windows larger than four seconds makes the estimate of the heart rate more robust but yields aggregated values over larger windows, which do not capture instantaneous heart rate very well.

B. Preprocessing

The Doppler audio was passed through a finite impulse response (FIR) antialiasing filter, then downsampled from 48 kHz to 1 kHz. The raw audio was then denoised using a standard wavelet filter implemented in MATLAB using sym4 wavelets and level $\lfloor \log_2 N \rfloor$ where N is the number of samples in the recording. The filtered audio was used for training and testing. We also investigate the use of the energy of the signals instead of raw signal as an input. The energy of the filtered audio over time was calculated using moving variance with a fixed window size of 125 ms. An illustration of the energy output is provided in Fig. 1.

		S1			S1-S2		
Input	Appr. (τ)	Prec	Rec	MAE	Prec	Rec	MAE
R	LSTM (70)	0.75	0.69	8.75	0.77	0.76	2.86
R	CNN (35)	0.87	0.75	2.19	0.87	0.85	3.11
E	LSTM (1)	N/A	N/A	N/A	0.80	0.80	14.92
E	CNN (70)	0.87	0.85	1.69	0.87	0.85	2.86

TABLE I: Performance for Fraction-of-Session (FS) Leave-out using Raw Signal (R) and its Energy (E). Top performances are colored blue. N/A values indicate that the methodology only converged to a trivial solution predicting a constant value.

C. Model Specification

Long Short-Term Memory (LSTM) Model. This model consists of: (1) an LSTM layer with 128 hidden units to capture the temporal information from the provided sequences, (2) a fully connected layer with ReLu activations and a 0.9 dropout rate, and (3) a final fully connected layer with softmax activations to perform the classification.

Convolutional Neural Network (CNN) Model. A simple 1D CNN model was implemented in this study. This model has two 1D convolutional layers with 64 filters and a kernel size of 3. A max-pooling layer is added to reduce the features learned by convolutional layers to 1/4 their size, and these features are flatten to a vector by a flatten layer. Since the CNN model learns features quickly, a dropout layer with 0.9 dropout rate is used to slow down the learning process and avoid overfitting. After the dropout process, the features are put into a fully connected layer with ReLu activation followed by another fully connected layer with Softmax activation to return the final prediction.

Post-Processing. The detections from the model may include some sporadic switches between different labels (e.g., see Fig. 2 [Top]). In order to enhance the detections, we apply a median filter. We select an optimal filter size τ based on the MAE of the estimated heart rate on the training data by first removing some outliers as discussed in Section III-A.

IV. Results and Discussion

A. Detection Performance on Doppler Data

For our analysis, we compare a number of scenarios by: (1) considering the use of S1 or the convex-hull of S1 and S2 as our target regions, (2) comparing LSTM and CNN models, (3) considering either the raw signal or the energy of the

		S1			S1-S2		
Input	Appr. (τ)	Prec	Rec	MAE	Prec	Rec	MAE
R	LSTM (100)	0.90	0.69	9.13	0.85	0.83	1.12
R	CNN (200)	0.92	0.78	1.67	0.83	0.86	1.11
E	LSTM (100)	0.88	0.90	1.40	0.88	0.90	1.40
E	CNN (200)	0.87	0.93	0.76	0.83	0.94	0.83

TABLE II: Performance for Entire-Session (ES) Leave-out using Raw Signal (R) and its Energy (E). Top performances are colored blue.

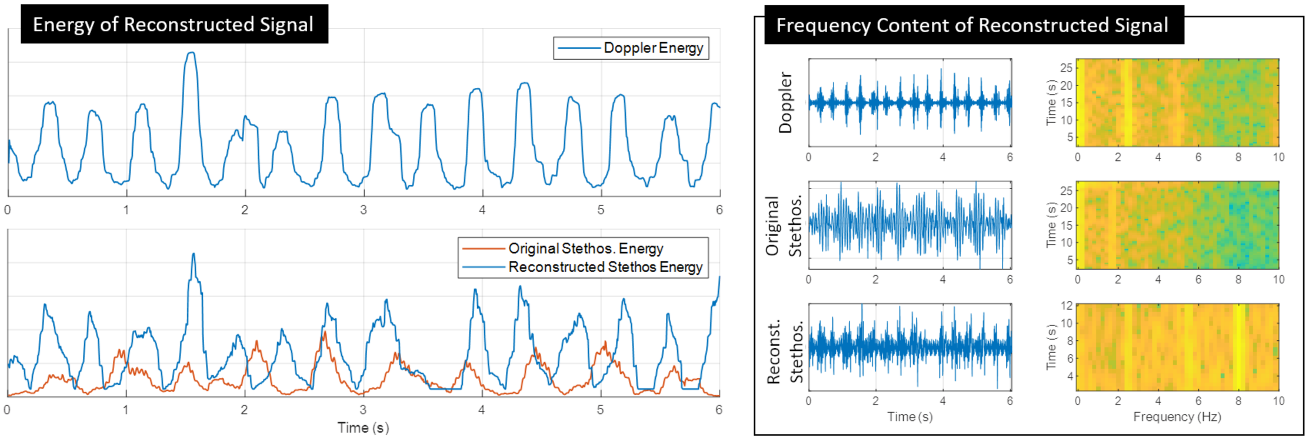

Fig. 3: Illustration of reconstructed stethoscope signal by matching the energy between stethoscope and Doppler using a CNN Network. [Left] The energy of the original and reconstructed stethoscope signals are compared to the energy of the Doppler. [Right] A snippet of the signals and corresponding spectrograms. We note that the dominant frequency component in the original stethoscope signal is different from the Doppler. The CNN corrects this content in the reconstructed signal.

signal as an input, and (4) considering a Fraction-of-Session (FS) or Entire-Sessions (ES) leave-out evaluation strategies. Tables I and II show the results. Overall, we identified the best performances to be obtained when using the energy of the signal as an input combined with a CNN model. Given our relatively small dataset and clean Doppler signals, it is logical that the energy is a reliable hand-crafted feature for heartbeats detection. The CNN architecture also takes advantage of temporal features in a more explicit form, so their higher performance also is expected. Given that we are provided more data for training, we hypothesize that the raw signal models would match (and possibly surpass) the energy-based approaches. We also observed that the detection of S1 on its own yields more accurate results. This is also expected since, upon visual inspection, the waveforms seem to be readily identifiable adding the S2 portion to the target region may just increase variability in the annotations because the S2 waveforms are harder to detect.

The problem of detecting heartbeats on entire new sessions, i.e., the ES Leave-out approach, should be more difficult than using fractions of the session for testing. However, this is not what was observed in Tables I and II. Upon closer inspection, one session had low signal quality which made it harder to predict. When using the FS Leave-out approach, we have a portion of this session included for testing; hence, the result is relatively higher error. However, for the ES Leave-out approach, this problematic session was used for training and consequently overall performance during testing improved.

B. Performance on Stethoscope Data

Unfortunately, when applying the various models to the stethoscope data, the models converged to trivial solutions predicting only a constant value (i.e., all zeros or ones). The quality of the stethoscope data was lower with significant background noise and lower frequency content. A possible cause of this problem is misalignment in the data. As observed in Fig. 1, the peaks in energy seem to align better with the S2 waveforms. These delays could be caused by the processing and buffering of the signals on the different systems, and it is not clear if these delays would remain constant across sessions. Furthermore, as it is also observed in Fig. 1, the standard filtering approaches did not appear to have meaningful impact on the stethoscope signals. We plan to explore these misalignment issues as well data-driven methodologies for filtering the signals in the future.

C. Extensions to Stethoscope Signal Reconstruction

As previously mentioned, data artifacts, such as background noise, pose a challenge to training a model for stethoscope signals. Hence, training a detector using the stethoscope signals requires more training data in order to capture all signal variability. Depending on the type of environmental sounds, this may translate into hundreds of hours of expert annotation. However, if we are able to record Doppler and stethoscope signals synchronously then we may use the Doppler signals information to train a classifier for the stethoscope. For example, we could train a detector for the Doppler and use the detection output on unlabelled data as input for training a model for the stethoscope. However, doing this. i) would require a reliable heartbeat detector for Doppler data and any changes to the detector would require recreation of all labels and ii) would not produce any interpretable results from the stethoscope (i.e., only detections would be produced that could not be verified in any way).

Instead, we propose reconstructing the stethoscope signal so it has some of the characteristics of the Doppler signal. Making audio signals match can be challenging due to the phase information of the signal. Hence, we propose focusing primarily on different content (e.g., the energy of the signal). We achieve that by defining a CNN that transforms the original stethoscope signal while trying to match their energy. The CNN consists of two 1D convolutional layers and each one is followed by a max-pooling layer. A dense layer is the last layer before the output. The loss function is constructed by the mean square error between Doppler energy and the reconstructed stethoscope energy. Fig. 3 illustrates the energy signals of the stethoscope before and after reconstruction, and how this gives us waveforms that are closer in shape to the Doppler signals and possess similar frequency content. We realize that an alternative to matching the energy is to directly reconstruct the magnitude of the spectrogram, which is something that we will explore in our future work. The advantage of this approach is that we can train this reconstruction network without having to rely on any labels, and it provides a reconstructed signal that is interpretable (i.e., it has similar frequency information as the Doppler and hence it can be played back to verify the locations of the heartbeats).

V. Conclusion and Future Work

This paper establishes preliminary work towards improving antenatal care through the use of Doppler and stethoscope devices that do not require expert placement for monitoring fetal cardiovascular health. We establish a baseline for Doppler audio-based heartbeat detection using the raw signal and the energy of the signal. In these cases, the energy of the signal yields the best results, detecting the heart rate with MAE as low to 0.76 beats per minute. Furthermore, we find that data collected from a single stethoscope recording device yields inconsistent and trivial solutions for the models considered. This is possibly due to the delays associated with sensor placement, the physics of sound traveling through the human body, and the delays introduced by the hardware used. Observing these issues in stethoscope data is expected given that the goal is to allow people to be able to monitor their child's health without requiring in-person expert supervision. Having established that baseline, we used CNNs to filter and reconstruct stethoscope data while also matching its energy to its Doppler energy counterpart, enhancing features useful for detection. In future work, we will expand on this effort to develop a stethoscope-based heartbeat detection pipeline; hence, extending crucial obstetric care outside of the clinic.

References

[1] M. F. MacDorman and E. C. Gregory, "Fetal and perinatal mortality: United states, 2013," *Natl Vital Stat Rep*, vol. 64, no. 8, pp. 1–24, 2015.

[2] R. Gagnon, K. Campbell, *et al.*, "Patterns of human fetal heart rate accelerations from 26 weeks to term," *Am J Obstet Gynecol*, vol. 157, no. 3, pp. 743–8, 1987.

[3] A. Samueloff, O. Langer, *et al.*, "Is fetal heart rate variability a good predictor of fetal outcome?" *Acta Obstet Gynecol Scand*, vol. 73, no. 1, pp. 39–44, 1994.

[4] U. Schneider, F. Bode, *et al.*, "Developmental milestones of the autonomic nervous system revealed via longitudinal monitoring of fetal heart rate variability," *PLoS One*, vol. 13, no. 7, e0200799, 2018.

[5] U. Schneider, B. Frank, *et al.*, "Human fetal heart rate variability-characteristics of autonomic regulation in the third trimester of gestation," *Journal of perinatal medicine*, vol. 36, no. 5, pp. 433–441, 2008.

[6] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, pp. 258–258, 2017.

[7] M. G. Signorini, A. Fanelli, *et al.*, "Monitoring fetal heart rate during pregnancy: Contributions from advanced signal processing and wearable technology," *Computational and Mathematical Methods in Medicine*, vol. 2014, p. 707 581, 2014.

[8] P. Van Leeuwen, L. Werner, *et al.*, "Fetal electrocardiographic measurements in the assessment of fetal heart rate variability in the antepartum period," *Physiol Meas*, vol. 35, no. 3, pp. 441–54, 2014.

[9] R. L. Goldenberg and E. M. McClure, "Maternal, fetal and neonatal mortality: Lessons learned from historical changes in high income countries and their potential application to low-income countries," *Matern Health Neonatol Perinatol*, vol. 1, 2015.

[10] D. L. Hoyert and E. C. Gregory, "Cause of fetal death: Data from the fetal death report, 2014," *Natl Vital Stat Rep*, vol. 65, no. 7, pp. 1–25, 2016.

[11] C. S. Bobrow and P. W. Soothill, "Causes and consequences of fetal acidosis," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 80, no. 3, F246, 1999.

[12] E. Gelson, R. Curry, *et al.*, "Effect of maternal heart disease on fetal growth," *Obstet Gynecol*, vol. 117, no. 4, pp. 886–91, 2011.

[13] M. Salameh, J. Lee, *et al.*, "Snoring and markers of fetal and placental wellbeing," *Clinica Chimica Acta*, vol. 485, pp. 139–143, 2018.

[14] J. Benjamin Emelia, P. Muntner, *et al.*, "Heart disease and stroke statistics—2019 update: A report from the american heart association," *Circulation*, vol. 139, no. 10, e56–e528, 2019.

[15] D. Gill, N. Gavrieli, *et al.*, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," in *Computers in Cardiology, 2005*, IEEE, 2005, pp. 957–960.

[16] D. B. Springer, L. Tarassenko, *et al.*, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2015.

[17] W. Zhang, J. Han, *et al.*, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomedical Signal Processing and Control*, vol. 32, pp. 20–28, 2017.

[18] M. Samieinasab and R. Sameni, "Fetal phonocardiogram extraction using single channel blind source separation," in *2015 23rd Iranian Conference on Electrical Engineering*, IEEE, 2015, pp. 78–83.

[19] S. Tomassini, A. Strazza, *et al.*, "Wavelet filtering of fetal phonocardiography: A comparative analysis," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 6034–6046, 2019.