FHSU-NETR: Transformer-based deep learning model for the detection of fetal heart sounds in phonocardiography

Murad Almadani[1], Mohanad Alkhodari[1,2], Samit Kumar Ghosh[1], Ahsan H. Khandoker[1]

[1] Healthcare Engineering Innovation Center (HEIC), Department of Biomedical Engineering, Khalifa University, Abu Dhabi, UAE

[2] Cardiovascular Clinical Research Facility, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Abstract

Assessing fetal well-being using conventional tools requires skilled clinicians for interpretation and can be susceptible to noise interference, especially during lengthy recordings or when maternal effects contaminate the signals. In this study, we present a novel transformer-based deep learning model called fetal heart sounds U-Net Transformer (FHSU-NETR) for automated extraction of fetal heart activity from raw phonocardiography (PCG) signals. The model was trained using a realistic synthetic dataset and validated on data recorded from 20 healthy mothers at the pregnancy outpatient clinic of Tohoku University Hospital, Japan. The model successfully extracted fetal PCG signals; achieving a heart rate mean difference of 2.72 bpm compared to the ground-truth calculated from fetal electrocardiogram (ECG). By leveraging deep learning, FHSU-NETR would facilitates timely interpretation of lengthy PCG recordings while reducing the heavy reliance on medical experts; thereby enhancing the efficiency in clinical practice.

1. Introduction

One of the most commonly used techniques to evaluate fetal well-being is phonocardiography (PCG); which is a non-invasive auscultation tool that enables continuous monitoring of fetal heart activity. PCG records the sounds generated from changes in blood flow or the opening and closing of heart valves [1].

Several signal processing techniques have been applied to extract fetal heart sounds from PCG recordings including kalman filtering [2], wavelet transform [3], empirical mode decomposition (EMD) [4], and blind source separation (BSS) [5]. While many of these techniques have demonstrated success in separating fetal heart sounds, they often require a reference fetal signal or prior estimation of maternal heart sounds for accurate performance [6]. Hence, there is a need for a more generalized technique that can overcome these limitations with minimal resources and without prior knowledge about the surrounding environment. This will enhance the reliability of fetal well-being assessment applications that utilize PCG.

In this paper, we propose, for the first time, the use of deep learning to extract fetal heart sounds from raw PCG recordings. To the best of our knowledge, deep learning as a tool for separating fetal heart sounds has not been investigated previously in the literature. Recently, U-Net models based on transformer neural networks have been widely used for medical image segmentation [7]. Our model, named fetal heart sounds U-Net Transformer (FHSU-NETR), is specifically designed to extract fetal heart sounds. It offers several advantages over conventional source separation approaches. It operates directly on raw PCG recordings without the need for prior source analysis to mitigate maternal sound effects. Furthermore, it eliminates the need for filtration or signal adjustment steps commonly used in previous studies. Additionally, FHSU-NETR learning from large datasets to make extraction decisions, rather than relying on single-input mathematical calculations.

2. Methodology

2.1. PCG Dataset Collection

The dataset utilized in this study consisted of 20 healthy pregnant women who were enrolled at the pregnancy outpatient clinic of Tohoku University Hospital in Japan [5]. Each patient was instructed to record a 10-minute PCG using a four-channel piezoelectric vibration system integrated into a high-definition three-dimensional (3D) printed plastic harness [5]. The harness was positioned on the abdomen; specifically above the belly button with each channel equidistant from that point. The recorded signals were amplified using a Powerlab 26T system with a resolu-

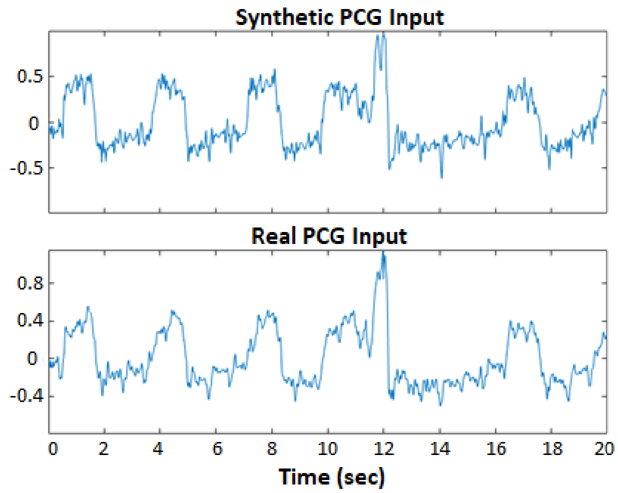

Figure 1. An example of the generated synthetic PCG versus the collected real PCG achieving 0.1 mean error.

tion of 16 bits and a sampling frequency of $1,000$ Hz. The acquisition protocol was approved by the Human Research Ethics Committee of Tohoku University Institutional Review Board (IRB), Sendai, Japan, and was conducted according to the declaration of Helsinki.

Simultaneously with PCG recordings, noninvasive electrocardiography (ECG) signals were also captured using IRIS, Atom Medical Co., Japan. Ten electrodes were placed on the maternal abdomen; with one reference electrode positioned on the back and one electrode located on the right thoracic region. Through a combination of cancellation techniques and reference blind source separation (BSS), a fetal ECG signal was successfully extracted [5].

2.2. Synthetic Data Generation

During data collection, the input PCG signal and fetal ECG (fECG) were obtained without ground-truth data. Our aim was to extract fetal PCG (fPCG) from raw PCG using supervised deep learning. However, substantial ground-truth data was essential for effective network training. To meet this requirement, synthetic PCG data containing ground-truth for fPCG, maternal PCG (mPCG), and maternal breathing signals were generated. Using a signal separation method from [6], the collected raw PCG signals were split into components. Although the resulting separated fPCG signal was suboptimal, the collected fECG guided placement of 10 different beat shapes extracted from [6]. These shapes comprehensively covered fetal heartbeats in the raw PCG. Thus, the trained network could identify and locate these beats during inference.

Merging the three signal samples created around $380,000$ synthetic raw PCG samples. Accompanied by corresponding ground-truth fetal and maternal heart sounds and breathing signals, these samples were used to train the deep learning model. Figure 1 compares synthetic maternal abdominal PCG (maPCG) with real data yielding an average root mean square error (RMSE) of approximately 0.1. This close-match affirms the model's training with realistic data; ensuring accurate performance representation.

2.3. Deep learning framework

The proposed pipeline is presented in Figure 2. Three one-dimensional U-Nets [7] that are linked to one another make up the pipeline. We used the right one for fetal heart sound, the upper-left one for maternal heart sound, and the lower-left one for maternal breathing extractions. The fetal heart sound and the maternal heart sound and breathing signals features were obtained by the three transformer-decoding procedures. The red lines show how the fetal heart sound features may be successfully retrieved from the maternal abdominal PCG sensors by eliminating the maternal heart sound and breathing features at each encoder step. We used the *tanh* activation after subtraction.

3. Results

The performance of the model was assessed by measuring the fetal heart rate in both 20-second intervals and the entire 10-minute recording of the masked fetal PCG segments. This evaluation involved comparing the results with the original fetal heart rate obtained from the ground-truth ECG mask. The heart rate was determined by counting the detected peaks in the fPCG signal; which were identified based on the envelope formed above the sound waves. Furthermore, a subjective comparison was conducted between the ECG-based and deep learning-based extracted fetal PCG signals; their performance was compared to the method proposed in [6] using two cases in their study (high and low wavelet threshold). An illustrative example of a 20-second segment of the extracted fetal PCG for a single patient is depicted in Figure 3.

This example highlights the first 7 peaks; demonstrating the effectiveness of the proposed model in capturing the fetal heart peaks. These peaks closely align with the fECG peaks; albeit with a slight and inconsistent delay which aligns with findings from recent studies (e.g., [5, 8]). The figure also showcases the superior performance of the proposed FHSU-NETR compared to the method proposed in [6]; as the latter fails to detect numerous fetal heart peaks that are accurately identified by the FHSU-NETR. This is noted with red lines in the Figure 3. Indicating a negligible discrepancy, the mean difference between the detected heart rate and the fECG heart rate across all patients amounted to -2.72, which was also close to the performance of previous studies summarized in Table 1.

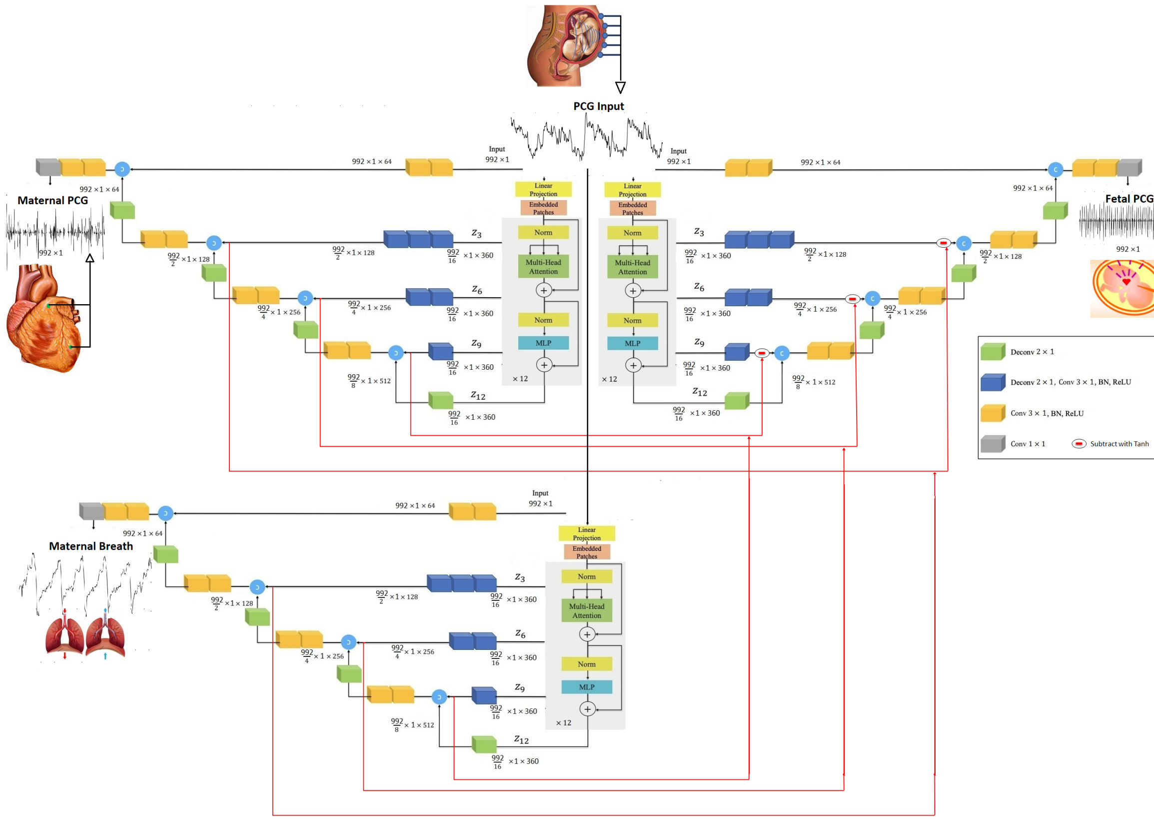

Figure 2. The proposed transformer-based fetal heart sounds U-Net Transformer (FHSU-NETR) pipeline for fetal heart sound extraction from abdominal phonocardiography. The pipeline consists of three UNetRs. Right for fetal heart sound, upper-left for maternal heart sound, and lower-left for maternal breath signal. The red lines depict the subtraction of the maternal breathing and heart sound features from the fetal heart sound (right) UNetR side at each transformer encoder step.

Table 1. Summary of studies that used phonocardiography (PCG) to estimate fetal heart rate.

Study	Year	# Participants	Data type	Pre-processing	Extraction method	Performance (relative to ground-truth)
Martinek *et al.* [9]	2017	10	PCG	Maternal removal linear/adaptive filter	Least mean squares	Mean difference: 1.85 bpm
Ibrahim *et al.* [6]	2017	20	PCG	Multi-resolution wavelet-based filter (WTST-NST)	Blind source separation	Mean difference: 0.21 bpm
Khandoker *et al.* [5]	2018	15	PCG	Bandstop filter Multi-resolution wavelet-based filter (WTST-NST)	Blind source separation	Mean difference: 1.30 bpm
Tomassini *et al.* [10]	2020	109	PCG	Bandpass filter Wavelet filter	Scalogram computations (AdvFPCG-Delineator)	Mean difference: 1.00 bpm
Ours	**2023**	**20**	**PCG**	**NA – Raw signals**	**Deep learning (FHSU-NETR)**	**Mean difference: 2.72 bpm**

4. Discussion and Conclusions

The proposed FHSU-NETR introduces a deep learning-based approach for accurate extraction of fetal heart sounds from raw PCG signals. The model was trained on a large dataset of more realistic synthetic inputs. This advancement enables the integration of automated artificial intelligence tools in clinical settings; reducing reliance on highly-experienced clinicians for the interpretation of lengthy, noisy, and mixed PCG signals. The performance evaluation of the model reveals minimal differences compared to the ground-truth fetal heart sounds (Figure 3); indicating its high overall performance. The low error in heart rate estimation achieved by FHSU-NETR demon-

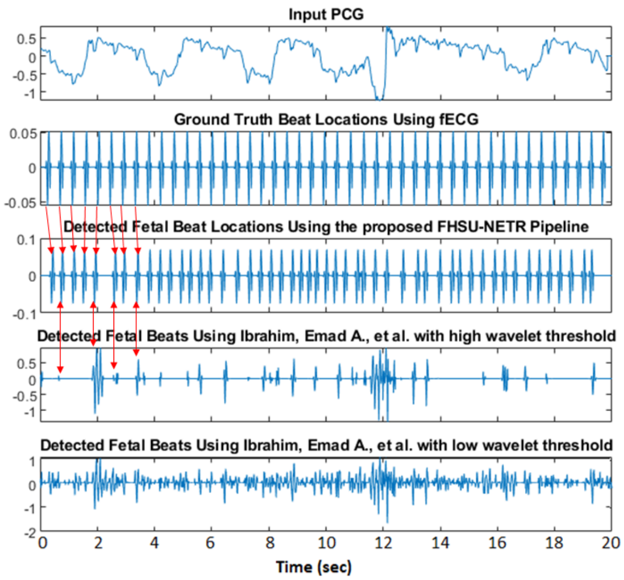

Figure 3. Example of a 20-second segment of the extracted fetal PCG. The red lines highlight the first seven fetal heart peaks detected by the FHSU-NETR.

strates its potential to act as a PCG-based technique for continuous evaluation of fetal health during pregnancy.

Compared to previous studies in the literature (Table 1), the model exhibits slightly higher differences from the ground-truth due to several reasons. Firstly, this study utilizes a challenging dataset; which comes in contrast to the works of [10] and [9] that evaluate their models on simpler datasets and synthetic data, respectively. Secondly, the methods listed in Table 1 often rely on multiple tuning parameters; reporting results based on the best parameter setting for each specific PCG input. This complexity hinders the real-world implementation of their models, whereas the proposed FHSU-NETR offers a straightforward and efficient implementation without the need for pre-adjustment processes. This distinction is evident in the last two plots of Figure 3a which illustrate the results obtained using the method proposed in [6] with two different levels of Wavelet thresholds. It is observed that a high threshold level produces a cleaner signal but at the expense of missing most fetal heart beats, while a low threshold level detects more beats but also introduces a high level of false detection due to retained noise.

Overall, the FHSU-NETR model, with its easy implementation, successfully extracts a greater number of fetal peaks compared to the original approach and with an acceptable margin of error. Furthermore, although other works suggest lower errors, they rely on mathematical formulations that may introduce biases rather than utilizing the superior generalizability of deep learning.

Acknowledgments

This work was supported by the Healthcare Engineering Innovation Center (HEIC) funding program, Khalifa University, Abu Dhabi, UAE, under Award 8474000132.

References

[1] Chourasia V, Tiwari A, Gangopadhyay R. A novel approach for phonocardiographic signals processing to make possible fetal heart rate evaluations. Digital Signal Processing 2014; 30:165–183.

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

[3] Koutsiana E, Hadjileontiadis L, Chouvarda I, Khandoker A. Fetal heart sounds detection using wavelet transform and fractal dimension. Frontiers in bioengineering and biotechnology 2017;5:49.

[4] Taralunga DD, Ungureanu M, Hurezeanu B, Strungaru R. Fetal heart rate estimation from phonocardiograms using an emd based method. In Proceedings of the 19th International Conference on Computers, Recent Advances in Computer Science (Zakynthos Island: Deauville). 2015; 414–417.

[5] Khandoker A, Ibrahim E, Oshio S, Kimura Y. Validation of beat by beat fetal heart signals acquired from four-channel fetal phonocardiogram with fetal electrocardiogram in healthy late pregnancy. Scientific reports 2018; 8(1):13635.

[6] Ibrahim EA, Al Awar S, Balayah ZH, Hadjileontiadis LJ, Khandoker AH. A comparative study on fetal heart rates estimated from fetal phonography and cardiotocography. Frontiers in physiology 2017;8:764.

[7] Hatamizadeh A, Tang Y, Nath V, et al. Unetr: Transformers for 3d medical image segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022; 574–584.

[8] Noorzadeh S. Extraction of fetal ECG and its characteristics using multi-modality. Ph.D. thesis, Université Grenoble Alpes, 2015.

[9] Martinek R, Nedoma J, Fajkus M, and others. A phonocardiographic-based fiber-optic sensor and adaptive filtering system for noninvasive continuous fetal heart rate monitoring. Sensors 2017;17(4):890.

[10] Tomassini S, Sbrollini A, Strazza A, and others. Advfpcg-delineator: Advanced delineator for fetal phonocardiography. Biomedical Signal Processing and Control 2020; 61:102021.

Address for correspondence:

Murad Almadani
Healthcare Engineering Innovation Center (HEIC), Department of Biomedical Engineering Khalifa University 127788, Abu Dhabi, UAE
100060577@ku.ac.ae, murad.almadani@gmail.com