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] 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 -1.5 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

Phonocardiography (PCG) is a widely used technique for fetal well-being evaluation, offering non-invasive continuous monitoring of fetal heart activity by recording sounds produced from blood flow changes and heart valve motions [1].

Various signal processing techniques, including Kalman filtering [2], wavelet transform [3], empirical mode decomposition (EMD) [4], and blind source separation (BSS) [5], have been employed to extract fetal heart sounds from PCG recordings. While successful, many of these methods rely on a reference fetal signal or prior maternal heart sound estimation for accuracy. Consequently, a more versatile technique is needed to overcome these limitations, offering enhanced reliability for fetal well-being assessment using PCG, all without the need for extensive resources or prior environmental knowledge.

This paper introduces a pioneering approach by harnessing deep learning to extract fetal heart sounds directly from raw PCG recordings. To the best of our knowledge, deep learning has not been previously explored for separating fetal heart sounds. Our model, named Fetal Heart Sounds U-Net Transformer (FHSU-NETR), leverages the transformer neural network-based U-Net model, which has gained popularity for medical image segmentation [6]. FHSU-NETR is purpose-built for fetal heart sound extraction and offers numerous advantages over conventional source separation methods. It operates directly on raw PCG recordings, bypassing the need for prior source analysis to mitigate maternal sound effects. Moreover, it eliminates the requirement for filtration or signal adjustment steps often seen in prior studies. FHSU-NETR's ability to learn from extensive datasets allows it to make extraction decisions, differing from single-input mathematical calculations in prior methods.

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 resolution of 16 bits and a sampling frequency of $1,000$ Hz.

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

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2023.026

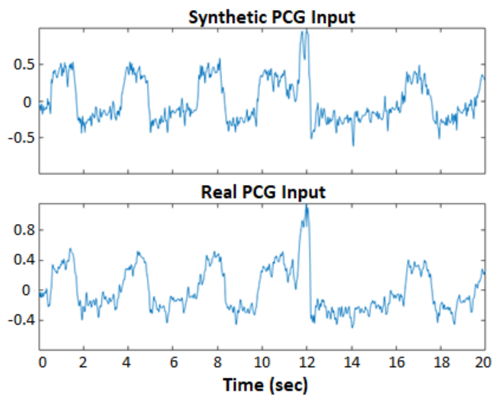

Figure 1. An example of the generated synthetic PCG versus the collected real PCG achieving 0.1 mean error.

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 [5], 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 [5]. 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 representation.

2.3. Deep learning framework

The proposed pipeline is presented in Figure 2. Three one-dimensional U-Nets [6] 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 model's performance 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 extracted fetal and maternal PCG signals, which were identified based on the envelope formed above the sound waves. Additionally, we show a comparison between the ECG-based and deep learning-based extracted fetal PCG signals, and their performance was benchmarked against the method proposed in [5] using two cases in their study (high and low wavelet threshold).

Figure 3a illustrates a 20-second segment of the extracted fetal PCG for a single patient in the dataset as an example. This example highlights the first 7 peaks, demonstrating the effectiveness of the proposed model in capturing fetal heart peaks. These peaks closely align with the fetal ECG peaks, albeit with a slight and inconsistent delay, in line with findings from recent studies (e.g., [5, 7]). For the same patient, the average fetal heart rate estimation was computed using 10-second segments of the extracted fPCG signal and the ground truth fECG, as demonstrated in Figure 3b. In this specific instance, the overall average heart rate difference between the detected heart rate and the fECG heart rate was a negligible -1.5, with a standard deviation of 18.13.

To visually compare the performance of our proposed method with the benchmark method in [5], Figure 3a underscores the superior capabilities of FHSU-NETR. It successfully detects numerous fetal heartbeats that the benchmark method struggles to identify, as evident from the red lines in the figure.

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

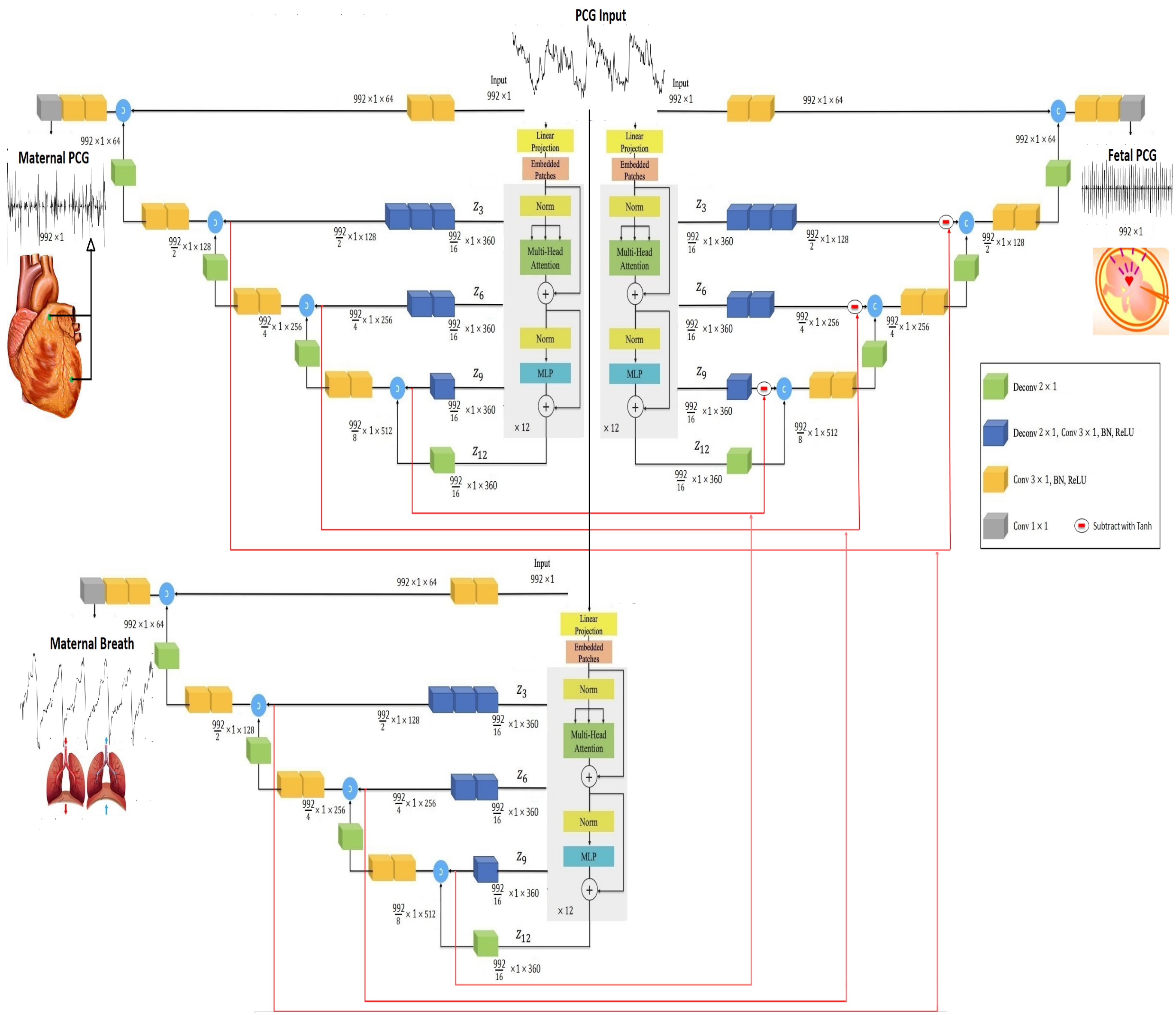

Figure 2. The proposed transformer-based fetal heart sounds U-Net Transformer (FHSU-NETR) pipeline for fetal heart sound extraction from abdominal phonocardiography. The pipeline consists of three UNetRs. Right for fetal heart sound, upper-left for maternal heart sound, and lower-left for maternal breath 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.

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

Furthermore, the proposed method was compared with the approach put forth in [5], which serves as a benchmark for fetal heart rate detection. the benchmark method relies on a subjective tuning parameter, increasing implementation complexity. This complexity hampers the practical real-world application of their models. In contrast, the proposed FHSU-NETR offers a straightforward and efficient implementation without the need for pre-adjustment processes. This distinction becomes evident in the last two plots of Figure 3a, which depict the outcomes obtained using the benchmark method proposed in [5] with two different levels of Wavelet thresholds. It is observed that a high threshold level produces a cleaner signal but at the

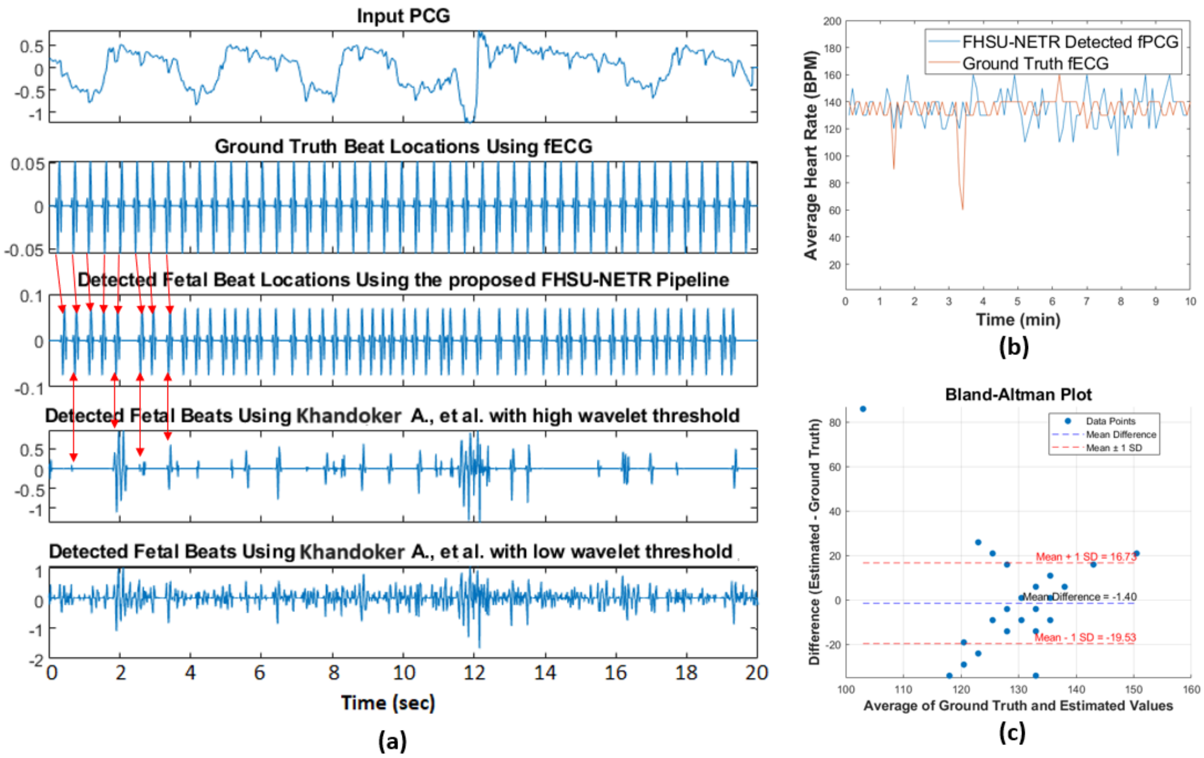

Figure 3. (a) Example of a 20-second segment of the extracted fetal PCG, The red lines highlight the first seven fetal heart peaks detected by the FHSU-NETR. (b) Average fetal heart rate estimation for the example in (a) across all recorded 10-minutes. (c) The corresponding Bland-altman plot.

cost of missing most fetal heartbeats. Conversely, a low threshold level detects more beats but introduces a high rate of false detection due to retained noise. This shows that the proposed method excels at identifying numerous fetal heartbeats that the benchmark method struggles to detect, as highlighted by the red lines in the figure. 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.

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] 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.

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

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