Attention in Fetal Peak R Detection and FECG Reconstruction

Gustavo R Faria[1], Vinicius C Rispoli[1], Gilmar S Beserra[1]

[1] University of Brasilia, Brasilia, Distrito Federal, Brasil

Abstract

The study aims to evaluate machine learning models for fetal ECG (fECG) reconstruction and R-peak detection from abdominal ECG (aECG) signals, focusing on the impact of input channel quantity on model accuracy. Three models were developed: Model I, a CNN-based AutoEncoder with 128 LSTM units and a self-attention module; Model II, a CNN-based AutoEncoder with 128 LSTM units and Bahdanau Attention; and Model III, a Transformer encoder with a CNN-based decoder. Using all four aECG input channels, Model I achieved an F1-score of 93.39%, Model II 96.71%, and Model III 96.06%. Model III had the lowest error rates (MSE: 0.0181, MAE: 0.0960). With a single aECG input channel, Model I's F1-score dropped to 88.86%, Model II to 90.04%, and Model III to 92.84%, with Model III again showing the lowest error rates. This suggests that Model III is better suited for fECG reconstruction and fetal R-peak detection, regardless of input channel quantity.

1. Introduction

Currently, several types of cardiac defects may occur during fetal heart development, with approximately 1 in every 125 newborns per year having some form of congenital heart malformation [1]. Regarding the etiology of congenital heart defects (CHDs), 85% to 90% of cases are reported to be attributed to external influences, while a minority are due to genetic factors. Moreover, the most severe defects commonly arise during the embryonic stage, particularly within the first eight weeks of gestation [2].

Therefore, prenatal monitoring is essential for early diagnosis, which enables improved care and increases the fetus's chances of survival during the postnatal period [3]. Several monitoring techniques are available, both invasive and non-invasive, including auscultation, cardiotocography (CTG), fetal electrocardiogram (fECG) using fetal scalp electrodes, and non-invasive fetal electrocardiogram (NI-fECG) [4]. The FHR (Fetal Heart Rate) can be obtained by analyzing the interval between consecutive R-peaks, where the R-peak is part of the QRS complex of the cardiac cycle, that indicates ventricular depolarization and atrial repolarization [5] [1].

Aiming to integrate the processing block of a system under development at FCTE-UnB for obtaining the FHR and fECG using only maternal abdominal ECG (aECG) signals, we studied several approaches for R-peak and fECG extraction, generally divided into two categories: Combined Sources (CS), which use electrodes placed on the maternal abdomen and thorax to subtract the maternal ECG (mECG) signal from the aECG in order to isolate the fECG; and Abdominal Electrodes Sourced (AES), which rely solely on the aECG signal [4]. Since the aECG signal is composed of mECG, fECG, and noise [4], this work aims to extract the fECG component from the aECG and detect the R-peaks to monitor the FHR.

The analysis of electrocardiograms has been increasingly optimized using artificial intelligence. In the context of CHDs, an example is the early detection of anomalies based on the NI-fECG signal and FHR during the prenatal period, which can even indicate the type of malformation present [6]. Artificial neural networks can be used to extract this information from the aECG [4]. Networks composed of convolutional layers are commonly used to extract relevant features from target signals [7]. In addition, recurrent neural networks (RNNs) are capable of processing temporal sequences and encoding information about the entire input signal. An implementation strategy involves the use of encoder–decoder models, which encode the input to retain essential information during processing and then decode it [8].

In this context, the objective of this work is to apply recurrent neural networks to detect and mark R-peaks and extract fECG from aECG. Consequently, the study is organized into four specific objectives: (1) demonstrate how effective the proposed models are at classifying R-peaks compared to those found in the literature; (2) enable the models to extract fetal R-peak features directly from the aECG, without using any mECG information; (3) compare the precision and fECG extraction capability when using four aECG channels as model input versus using only a single input channel; (4) use the same benchmark datasets commonly cited in the literature for model training and evaluation to ensure fair performance comparison.

2. Methods

In this study, three models based on recurrent neural networks were made for the detection of the fetal R-peak and the fECG extraction. Initially, a preprocessing step was applied to the signals to eliminate unwanted noise and reduce computational cost. Subsequently, performance metrics were defined for model evaluation, followed by the development of the final model architecture.

2.1. *Dataset*

The dataset used in this work was the A&D fECG (abdominal and direct fetal electrocardiogram) dataset [9], which consists of five signal sets collected from different pregnant women between the 38th and 41st weeks of gestation: *r01*, *r04*, *r07*, *r08*, and *r10*. Each set includes five channels: four corresponding to aECG signals and one to fECG, invasively collected via fetal scalp electrodes. Each recording has a duration of five minutes and a sampling frequency of 1 kHz. Additionally, each dataset includes manually annotated fetal R-peaks by clinical experts. During preprocessing, it was observed that the *r10* set contained segments without R-peak annotations (188s – 190s; 203s – 210s), which were therefore removed.

2.2. Preprocessing

To prepare the dataset, GNU Octave was used to apply a *Butterworth* band-pass filter to the aECG signals, with cutoff frequencies set between 1 Hz and 100 Hz, removing unwanted noise while preserving both mECG and fECG components [10]. Then, the sampling rate was downsampled to 200 Hz, and the signals were segmented into one-second sections to reduce the computational load during processing. Finally, the signals were normalized. Only segments containing at least one annotated fetal R-peak were selected and divided into 70% for training, 15% for validation, and 15% for testing. An example of an input aECG signal is four aECG channels and a fifth channel representing the invasively collected fECG used for training, with manually annotated R-peaks.

2.3. Metrics

During model training, the metric used for R-peak detection was accuracy, which measures the ratio of correct predictions to the total number of predictions. Another metric used for model evaluation was the F1-score, which provides the harmonic mean of precision and recall [11].

For a more comprehensive analysis in the test set, we used the F1-score, PPV, SE, and ACC metrics for R-peak detection, as well as MSE and MAE during validation and training of the fECG reconstruction.

Given that the signal duration is 1 s and the QRS complex typically lasts between 70 ms and 100 ms, a tolerance margin of 10% was applied to the F1-score and accuracy calculations on the test set 5% to the left and 5% to the right of the actual R-peak instant [12]. A threshold of 0.5 was applied to the model's outputs to determine whether a prediction was considered valid.

The binary cross-entropy loss function, which measures the discrepancy between predicted and actual labels in binary classification tasks [8], was used during model training and validation for R-peak labeling. And the loss function for the fECG reconstruction we used was MSE.

2.4. Model's architecture

Model-I consists of an attention-based autoencoder built with convolutional layers and LSTM units. The first stage is an encoder composed of four 1D convolutional layers, with filter counts and kernel sizes as follows: $32,\ 9 \times 9$; $32,\ 6 \times 6$; $64,\ 6 \times 6$; and $128,\ 3 \times 3$. Each layer applies L1 regularization to the filters to reduce the risk of overfitting. After each convolution, batch normalization layers are applied to normalize the outputs during training, thereby improving model convergence [8]. Subsequently, the ReLU non-linear activation function is applied. A max-pooling operation is used in the last three convolutional layers to halve the size of their outputs. In addition, 128 LSTM units are used to learn the temporal dependencies and features extracted by the encoder. The decoder is composed of three convolutional layers, also with L1 regularization, and kernel sizes of $128,\ 3 \times 3$; $64,\ 3 \times 3$; and $32,\ 3 \times 3$, respectively. The up-sampling function is applied three times to double the signal length at each step, restoring the original input dimensions. Batch normalization and ReLU activation are used in the decoder in the same way as in the encoder. After the signal is reconstructed by the decoder, a self-attention head is applied to extract the most relevant features from the reconstructed signal. The model outputs are generated by a final 1D convolutional layer, with a sigmoid activation function that provides the probability of an R-peak at each sample point and an ELU function for the fECG reconstruction.

Model-II features an architecture similar to Model-I, comprising the same encoder and decoder, and using 128 LSTM units. However, it replaces the self-attention head with a layer based on the Bahdanau Attention mechanism, positioned after the LSTM units and before the decoder. This layer generates a context vector from the LSTM output and its hidden states. The output layers of this model are the same as those in the first model.

Model-III consists of an encoder with two 1D convolutional layers, each using filters of size $128,\ 3 \times 3$. In addition, it includes a Transformer Encoder structure, comprising a Multi-Head Attention (MHA) mechanism followed by a normalization layer placed between the Transformer's input and the MHA output. Two feedforward dense layers are included: the first with 256 neurons and ReLU activation, and the second with 128 neurons. A second normalization layer is applied between the output of the first normalization layer and the output of the last dense layer. Subsequently, a CNN-based decoder is employed, featuring two convolutional layers with filters of size 128, 3×3 and 64, 3×3, along with Batch Normalization and ReLU activation applied between the layers, which are connected via an UpSampling function. The output layers are identical to those used in the previous models.

The three models were trained using the Adam optimizer with a learning rate of 5×10^{-4}. Training was performed over 150 epochs for every model with a batch size of 32, on *Google Colaboratory*.

3. Results and Discussion

We conducted a comparison between state-of-the-art approaches and the proposed models for fetal R-peak detection, as shown in Table 1. All methods considered used only a single input channel. The table presents the F1-score achieved by each model, along with the respective datasets employed: A&D FECG (Abdominal and Direct fECG) [9] and PCDB (PhysioNet Challenge Database) [13]. This shows that the proposed models are equivalent to state-of-the-art approaches.

Table 1. Comparison between the proposed models and the state of the art.

Model	F1-score	Dataset
Dual Attention AE [14]	$98,01\%$	A&D FECG
U-Net [15]	$93,01\%$	PCDB
Model-III	$92,84\%$	A&D FECG
SFTF-GAN [16]	$90,05\%$	A&D FECG
Model-II	$90,04\%$	A&D FECG
Model-I	$88,86\%$	A&D FECG

With confusion matrices, we computed the performance metrics presented in Tables 2 and 3. Based on these results, we compared the models with each other and assessed the impact of using multiple input channels versus a single input channel. From Table 2, Model-II achieved the best performance for R-peak detection when using four input channels, with an F1-score of 96.71%. However, Model-III achieved the lowest error rate in the reconstruction of the fECG signal, a MSE equal to 0.0181 with four aECGs as inputs. Model-I underperformed on all evaluation metrics, indicating that autoencoder-based models require more robust architectures, as performed in [14].

The R-peak detection comparison between all models is illustrated in the Figure 1 (e), (f) and (g). One can

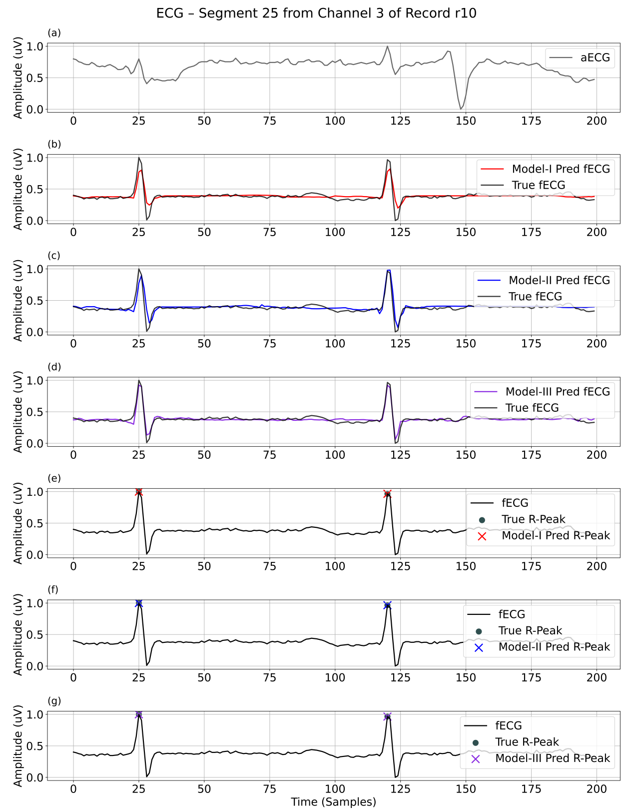

Figure 1. Exemple of a detection of fetal R-peaks and fECG reconstruction. (a) One channel aECG as input; (b) Model-I's predicted fECG; (c) Model-II's predicted fECG; (d) Model-III's predicted fECG; (e) Model-I's predicted R-peaks; (f) Model-II's predicted R-peaks; (g) Model-III's predicted R-peaks.

Table 2. Performance of the models when four aECG channels were used as input.

	Model-I	Model-II	Model-III
F1-score	93.39%	96.71%	96.06%
PPV	99.17%	98.69%	98.89%
SE	88.24%	94.80%	93.38%
ACC	97.30%	98.60%	98.34%
MSE	0.0217	0.0198	0.0181

see that all three models had good accuracy in detecting the R-peak. Furthermore, it is shown in Figure 1 (b), (c) and (d) the differences for reconstructing a fECG signal. The Table 3 shows that Model-III remained efficient in both detection and reconstruction tasks, indicating that Transformer-based architectures are capable of handling the task more effectively. However, Model-II showed a significant drop in its accuracy when using only one input channel, with a decrease of approximately 6% in F1-score and more than 11% in Sensitivity.

In the case of Model I, an unexpected behavior was observed, as the fECG reconstruction capacity did not decrease; on the contrary, a slight improvement was observed.

Table 3. Performance of the models when one aECG channel was used as input.

	Model-I	**Model-II**	**Model-III**
F1-score	88.86%	90.04%	92.84%
PPV	97.23%	98.00%	98.00%
SE	81.82%	83.27%	88.19%
ACC	95.51%	95.97%	97.02%
MSE	0.0214	0.0202	0.0193

4. Conclusion

Model-I performed below the other two models in both conditions, but was able to maintain its quality in fECG prediction even with only one aECG as input. Model-II, despite the highest F1-score with all channels, struggled with reduced input. Model-III consistently performed well in both reconstruction and detection tasks across input configurations, making it the most robust and comprehensive solution. Moreover, it contains a number of parameters comparable to the first and lower than the second. This suggests that Model-III is better suited for fECG reconstruction and detection of the fetal R peak, regardless of the input channel quantity, since its F1 score was only reduced 3.22%, while maintaining its MSE below 0.02. For future work, modifications should be made so that the model can be used in real-time applications.

Acknowledgments

This study was carried out with the support of the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES).

References

[1] Sameni R, Clifford GD. A review of fetal ecg signal processing issues and promising directions. Electrophysiology Therapy Journal 2010;.

[2] Zubrzycki M, Schramm R, Costard-Jäckle A, Grohmann J, Gummert JF, Zubrzycka M. Cardiac development and factors influencing the development of congenital heart defects (chds): Part i. International journal of molecular sciences 2024;.

[3] van Velzen CL, Haak MC, Reijnders G, Rijlaarsdam MEB, Bax CJ, Pajkrt E, Hruda J, Galindo-Garre F, Bilardo CM, de Groot CJM, Blom NA, Clur SA. Prenatal detection of transposition of the great arteries reduces mortality and morbidity. Ultrasound in Obstetrics Gynecology 2015;.

[4] Kahankova R, Martinek R, Jaros R, Behbehani K, Matonia A, Jezewski M, Behar JA. A review of signal processing techniques for non-invasive fetal electrocardiography, 2020.

[5] Smith V, Arunthavanathan S, Nair A, Ansermet D, da Silva Costa F, Wallace EM. A systematic review of cardiac time intervals utilising non-invasive fetal electrocardiogram in normal fetuses. BMC Pregnancy and Childbirth 2018;.

[6] de Vries IR, van Laar JOEH, van der Hout-van der Jagt MB, Clur SAB, Vullings R. Fetal electrocardiography and artificial intelligence for prenatal detection of congenital heart disease. Acta obstetricia et gynecologica Scandinavica 2023;.

[7] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. In Proceedings of 2010 IEEE International Symposium on Circuits and Systems. IEEE, 2010; .

[8] Prince SJD. Understanding Deep Learning. 2024.

[9] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark C. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals 2000;.

[10] Zhang N, Zhang J, Li H, Mumini O, Samuel O, Ivanov K, Wang L. A novel technique for fetal ecg extraction using single-channel abdominal recording. Sensors 2017;.

[11] Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. On evaluation metrics for medical applications of artificial intelligence. Scientific Reports 2022;.

[12] Darmawahyuni A, Tutuko B, Nurmaini S, Rachmatullah MN, Ardiansyah M, Firdaus F, Sapitri AI, Islami A. Accurate fetal qrs-complex classification from abdominal electrocardiogram using deep learning. International Journal of Computational Intelligence Systems 2023;.

[13] Silva I, Behar J, Sameni R, Zhu T, Oster J, Clifford GD, Moody GB. Noninvasive fetal ecg: the physionet/computing in cardiology challenge 2013 ;.

[14] Ghonchi H, Abolghasemi V. A dual attention-based autoencoder model for fetal ecg extraction from abdominal signals. IEEE Sensors Journal 2022;.

[15] Zhou P, Schwerin B, So S. U-net based fetal r-peak prediction from abdominal ecg signals. In 2024 9th International Conference on Signal and Image Processing. 2024; .

[16] Zhong W, Zhao W. Fetal ecg extraction using short time fourier transform and generative adversarial networks. Physiological Measurement 2021;.

Address for correspondence:

Gustavo Raspante Faria
University of Brasilia, Brasilia, Distrito Federal, Brasil
gustavo.faria@aluno.unb.br