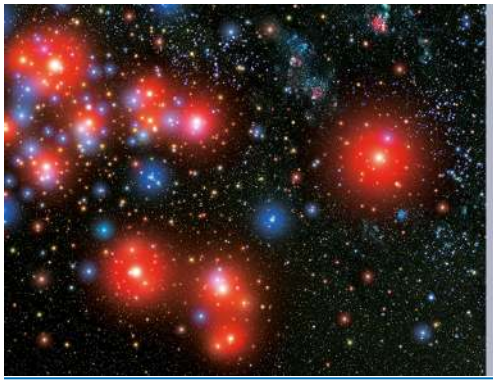
End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising



IOP Publishing Physiol. Meas. 41 (2020) 015005 (11pp) https://doi.org/10.1088/1361-6579/ab69b9

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising

Eleni Fotiadou[1], Tomasz Konopczyński[2], Jürgen Hesser[2] and Rik Vullings[1]

[1] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven 5612 AP, The Netherlands
[2] Department of Radiation Oncology, Medical Faculty Mannheim, Central Institute of Scientific Computing (IWR) and Central Institute for Computer Engineering (ZITI), Heidelberg University, Germany

E-mail: E.Fotiadou@tue.nl

RECEIVED
23 August 2019

REVISED
13 December 2019

ACCEPTED FOR PUBLICATION
9 January 2020

PUBLISHED
5 February 2020

Keywords: convolutional neural networks, encoder–decoder network, fetal ECG denoising, fetal ECG enhancement, fetal electrocardiography

Abstract

Objective: Non-invasive fetal electrocardiography has the potential to provide vital information for evaluating the health status of the fetus. However, the low signal-to-noise ratio of the fetal electrocardiogram (ECG) impedes the applicability of the method in clinical practice. Quality improvement of the fetal ECG is of great importance for providing accurate information to enable support in medical decision-making. In this paper we propose the use of artificial intelligence for the task of one-channel fetal ECG enhancement as a post-processing step after maternal ECG suppression. *Approach*: We propose a deep fully convolutional encoder–decoder framework, learning end-to-end mappings from noise-contaminated fetal ECGs to clean ones. Symmetric skip-layer connections are used between corresponding convolutional and transposed convolutional layers to help recover the signal details. *Main results*: Experiments on synthetic data show an average improvement of 7.5 dB in the signal-to-noise ratio (SNR) for input SNRs in the range of −15 to 15 dB. Application of the method with real signals and subsequent ECG interval analysis demonstrates a root mean square error of 9.9 and 14 ms for the PR and QT intervals, respectively, when compared with simultaneous scalp measurements. The proposed network can achieve substantial noise removal on both synthetic and real data. In cases of highly noise-contaminated signals some morphological features might be unreliably reconstructed. *Significance*: The presented method has the advantage of preserving individual variations in pulse shape and beat-to-beat intervals. Moreover, no prior knowledge on the power spectra of the noise or the pulse locations is required.

1. Introduction

During pregnancy and labor, monitoring of the condition of the fetal heart is of paramount importance. Fetal monitoring can support medical decision taking, while early disease diagnosis can increase the effectiveness of the appropriate treatment. Since a fetus is well protected within a woman's womb it is inaccessible for direct measurements. Nowadays, the gold standard for fetal heart assessment is cardiotocography, which provides a visual representation of the fetal heart rate together with uterine contractions (Signorini *et al* 2003). However, cardiotocography is prone to signal loss and does not provide the heart rate with beat-to-beat variations. A non-invasive fetal ECG, derived from abdominal electrodes, has the potential to provide beat-to-beat heart rate information (Jezewski *et al* 2006) with the added possibility of assessing ECG morphology such as the PR and QT intervals or ST segments. Despite the ease of its applicability, fetal ECG signals are substantially contaminated by interference and noise which vary depending on gestational age, the position of the electrodes, skin impedance etc (van Laar *et al* 2014). Most significantly, the signals are masked by the maternal ECG and background noise caused by the abdominal and uterine electromyogram. Additional noises, such as electrode movement and powerline interference, and the multiple layers of dielectric biological tissues that the signals must pass further lower the signal-to-noise ratio (SNR) of the fetal ECG. The high number of interfering sources that are typically

non-stationary and overlap with the fetal ECG in time and frequency domains render fetal ECG extraction and heart rate detection challenging signal processing tasks and limit the applicability of the method in clinical practice.

Despite the advances in adult ECG signal processing, analysis and interpretation of the fetal ECG is still in its infancy. Various techniques have been proposed in the literature for extracting the fetal ECG from non-invasive abdominal recordings. The main methods include adaptive filtering (Widrow *et al* 1975, Adam and Shavit 1990, Sameni 2008, Niknazar *et al* 2013), blind source separation (Martín-Clemente *et al* 2002, Ye *et al* 2009, Zhang *et al* 2009, Camargo-Olivares *et al* 2011) and template subtraction (Cerutti *et al* 1986, Ungureanu *et al* 2007, Vullings *et al* 2009). Clifford *et al* (2014) gave a thorough review of the key achievements and the follow-up research as a result of the PhysioNet/Computing in Cardiology Challenge 2013 (Silva *et al* 2013). The aim of the challenge was to encourage heart rate estimation and QT interval measurements in an automated manner, and it succeeded in stimulating research in these areas (Andreotti *et al* 2013, Behar *et al* 2013, Varanini *et al* 2013, Podziemski and Gieraltowski 2013). In general, a five-step approach is followed by most algorithms: this includes pre-processing, maternal ECG estimation and subtraction, fetal heart rate estimation and post-processing of heart rates. Usually, the extracted fetal ECG still has a low SNR and additional processing is required to further enhance its quality. One traditional approach for improving the SNR of the signal is beat-to-beat averaging (Lindecrantz 1983) at the expense of loss in the individual variations in pulse shape. In a previous work (Fotiadou *et al* 2018) the authors used a time-sequenced adaptive least mean squares (LMS) filter as a post-processing step to enhance the quality of the extracted multichannel fetal ECG. Despite substantial noise removal, the method requires knowledge of the location of the R peak prior to use and is not designed to handle cases of arrhythmia.

Recently, deep neural network models, such as stacked denoising autoencoders, recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have been widely used in the area of signal and image denoising, with great success (Jain and Seung 2008, Maas *et al* 2012, Lu *et al* 2013, Mao *et al* 2016, Zhang *et al* 2017). Several attempts have also been made in the field of ECG signal processing, for example adult ECG denoising (Xiong *et al* 2016, Antczak 2018), adult arrhythmia detection (Isin and Ozdalili 2017) and fetal QRS complex detection (Zhong *et al* 2018, Lee *et al* 2018). Xiong *et al* (2016) used an improved denoising autoencoder (DAE), reformed by a wavelet transform method to remove baseline wander noise, electrode contact noises and motion artifacts from adult ECG signals. Antczak (2018) trained a RNN model to de-noise synthetic ECG data and subsequently used transfer learning to enhance the quality of real data.

To the best of our knowledge there has been no work that has attempted fetal ECG denoising using CNNs. Inspired by the work of Mao *et al* (2016) on image restoration, we propose the use of a deep convolutional encoder–decoder network with symmetric skip-layer connections for single-channel fetal ECG denoising. Residual noise in the fetal ECG, after the maternal ECG has been removed, is often non-stationary, complex and has spectral overlap with the fetal ECG. Our method removes the residual noise by capturing the structure of the fetal ECG by the convolutional layers and recovering the signal details with the help of the transposed convolutional layers.

The rest of the paper is organized as follows. Section 2 presents the proposed fetal ECG enhancement method and the data used. Experimental results are provided in section 3. Finally, the results are discussed in section 4 and conclusions are drawn in section 5.

2. Methods and data

2.1. Network architecture

In this section, we present the proposed CNN model for removing the residual noise after extracting the fetal ECG from antenatal abdominal recordings. We adopt the network architecture of Mao *et al* (2016), developed for image restoration, and modify it to make it suitable for ECG signal denoising. The proposed method, as shown in figure 1, aims to learn an end-to-end noise removal function from noise-corrupted fetal ECG to its clean version. The network contains two stages: encoding and decoding. The encoder acts as a feature extractor that removes the noise while preserving the primary ECG components. The decoder recovers the signal details and delivers a 'clean' ECG as an output. Skip connections are added between every two corresponding convolutional and transposed convolutional layers to help with recovering a clean ECG while also tackling the optimization difficulty caused by gradient vanishing in deep architectures.

2.2. Network parameters

The network depth together with the selection of kernel size and the use or not of subsampling or dilation in the convolutional layers define the receptive field of the network. Increasing the size of the receptive field can make use of the context information in a larger signal region. It was indicated that the effective patch size of the denoising methods is highly correlated with the receptive field of the network (Jain and Seung 2008, Burger *et al* 2012). A simple strategy to achieve a large receptive field is to increase the number of layers in the network.

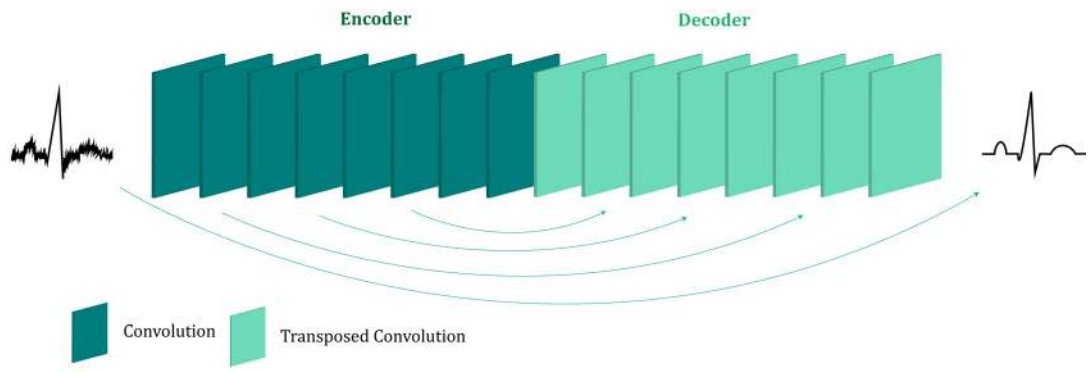


Figure 1. The architecture of the proposed fetal ECG denoising network. The network consists of an encoder of eight convolutional layers and a decoder of eight transposed convolutional layers. Symmetric skip connections are added between every two corresponding convolutional and transposed convolutional layers.

However, this becomes computationally very intensive. Alternatively, subsampling operations can be performed by the network, but this is not generally preferred in denoising tasks for the sake of preserving the signal details (Zhang *et al* 2017).

The fetal ECG usually exhibits high levels of noise and thus requires a large effective patch size to capture more information for efficient denoising. Moreover, in order to exploit the self-similarity of the underlying fetal ECG signal the network should permit correlations to extend to several heartbeats. In the proposed framework we decided to use a relatively deep network of eight convolutional and eight mirrored transposed convolutional layers. Both the convolutions and the transposed convolutions are one-dimensional. Furthermore, subsampling by two is employed after every convolutional layer apart from the first one. We found that including subsampling operations in the convolutional layers helps rather than hinders in terms of denoising performance, since the receptive field of the network is drastically increased. Without subsampling a huge number of convolutional layers would have to be added to obtain the same receptive field. Moreover, in practical implementation of the network, employment of the network should be fast and subsampling is also beneficial as it reduces the size of feature maps. It should be noted here that the additional skip connections account for the lost signal details introduced by the subsampling to a great extent. Regarding the kernel size, we use 15, which for the case of signals sampled with 500 Hz corresponds to 30 ms. With the above choices an effective receptive field of roughly 3.6 s is achieved that corresponds to 5–10 heartbeats. For the sake of computational efficiency we decided not to add more layers, even though this would increase the receptive field even further.

The input of the network is fixed at 1920 samples, slightly higher than the effective receptive field (1800 samples), to facilitate subsampling operations. The number of filters applied to produce the feature maps is [64, 128, 256, 256, 512, 512, 1024, 2048] for the eight convolutional layers, respectively. Consequently, considering that the encoder subsamples the input signal by two in seven layers, the dimension of the bottleneck feature vector is 2048×15. The transposed convolutional layers use a mirrored numbered of filters. Leaky rectified linear units are utilized for non-linearity after each layer. Implementation of the network was done in Keras.

2.3. Skip connections

In deep networks, transposed convolution does not work very well in recovering the details of the input data from just the data abstraction, possibly because too much detail has already been lost in the convolution (Mao *et al* 2016). The suggested network is not only deep but, as mentioned above, the encoder heavily downsamples the input signal. To address this problem, skip connections are added from every two convolutional layers to the corresponding mirrored transposed convolutional layers. The feature maps passed by the skip connections carry much signal detail, which helps transposed convolution to recover an improved clean version of the fetal ECG. In addition, skip connections facilitate training of deep networks as they aid in back-propagating the gradient to the bottom layers (Srivastava *et al* 2015, He *et al* 2016). The way in which the skip connections are used in the network is demonstrated in figure 2. As can been seen from the figure, the output of a convolutional layer is added to the output of the corresponding transposed convolutional layer and subsequently an activation function is applied to their sum.

2.4. Data

The fetal ECG signals, even after the maternal ECG has been removed, are still affected by noise. It is hence impossible to have real fetal ECG pairs (noisy and clean) to train the network. Thus, we used a rich dataset of simulated fetal ECG data for training. The algorithm was then extensively validated in a separate simulated test

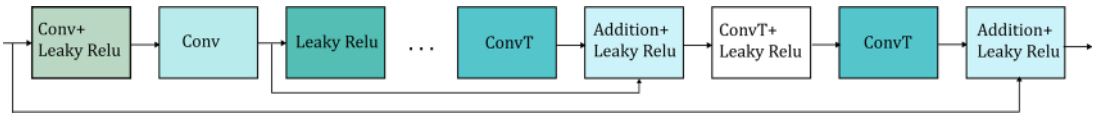


Figure 2. The way that the skip connections are used in the proposed network. Only the first two skip connections are shown for simplicity. Conv stands for convolution, ConvT for transposed convolution and Relu for rectified linear units.

dataset, while results are additionally provided for real signals and compared with simultaneously performed scalp fetal ECG recordings.

2.4.1. Simulated data

We used the fecgsyn toolbox (Andreotti *et al* 2016) to create an extensive simulated dataset. By employing fecgsyn fetal–maternal mixtures can be created while there is the possibility to model a number of non-stationary events that affect the morphology and dynamics of the abdominal ECG by rotating, translating and modulating the available sources. The synthetic abdominal signals consist of 34 channels (32 abdominal and two maternal ECG reference channels). With the use of the toolbox, we created a dataset where different physiological events are considered, such as ectopic beats, uterine contraction, noise, fetal movement etc, similar to the Fetal ECG Synthetic Database (FECGSYDB) (Andreotti *et al* 2016). Unfortunately, the data simulated with the fecgsyn toolbox is based on nine different vectorcardiograms (VCGs) only. This means that training merely on this dataset has the risk of overfitting to these specific VCGs since there is relatively small variation in the timing of the ECG intervals. One possible way to introduce more variety into the dataset is to model more VCGs, but we have chosen a different approach. Our approach includes modifying the fecgsyn toolbox to obtain a large variety of ECG morphologies based on the already available VCGs. The modified fecgsyn toolbox is able to receive a VCG and then apply random modifications of the lengths of the VCG intervals and segments, as well as of the amplitudes of the waves, such that a new VCG is created. This VCG was then used as a base to create the abdominal fetal ECG. The ranges of modifications were selected to cover a wide range of variations of the morphological features of the ECG while still ensuring physiologically plausible ECGs.

To further increase the diversity of ECG patterns in our dataset we generated one additional set of data. For this set, adult ECG signals were used to simulate the fetal ECG. Morphologically, adults and fetuses have rather similar ECG patterns: they both comprise a PQRST complex, yet the amplitudes and lengths of adult ECG segments and intervals are large compared with those of the fetal ECG. The adult ECGs were collected from the different Physionet databases (Goldberger *et al* 2000), the PTB Diagnostic ECG Database (Bousseljot *et al* 1995), the St Petersburg Institute of Cardiological Technics 12-lead Arrythmia Database (Goldberger *et al* 2000) and the QT database (Laguna *et al* 1997). The signals were pre-processed to remove baseline wander and noise and to resemble the fetal ECG. A high-pass filter with cut-off frequency at 1 Hz was first applied to the signals, followed by Savitzky–Golay filtering of order 8 and length 31. Based on the fact that the fetal heart rate is two to three times faster than the adult heart rate, the adult ECG signals were resampled at half frequency to look similar to fetal ECG signals. Amplitude scaling of the ECG was not necessary, as the data were normalized before entering the network. As a final step, noise was added to the signals. For this purpose, a number of six-channel abdominal recordings, for which the study protocol is described in Verdurmen *et al* (2016), were employed. In some recordings, after suppression of the maternal ECG, the fetal ECG was impossible to detect in any electrode or subset of them. This was caused either by shielding of the fetus by the vernix caseosa or because some electrodes were too far from the fetal heart. In these cases, we assumed that the electrodes recorded only noise (apart from the maternal ECG that is subtracted) and this noise was added to the pre-processed adult ECG to simulate the noisy fetal ECG. The advantage of using real noise recordings is that the data used for training the network are highly similar to the data we aim to use in testing and ultimately employ in the network. By learning how to remove real noise, the network is likely to perform better with real data.

2.4.2. Real data

To investigate the impact of training with synthetic data on denoising in a real dataset, we used the Abdominal and Direct Fetal Electrocardiogram Database (Jezewski *et al* 2012). This database contains multichannel fetal ECG recordings obtained from five different women in labor between 38 and 41 weeks of gestation. Each recording contains four signals acquired from the maternal abdomen and one scalp ECG signal. The recordings have a duration of 5 min and are sampled at 1000 Hz.

2.4.3. Data pre-processing

The signals of all datasets were pre-processed before entering the network either for training or for testing. Regarding the abdominal recordings, the open-source algorithm of Varanini *et al* (2013) was initially applied

to them. According to this algorithm, first the baseline wander and the powerline interference were removed. After that, the maternal ECG was estimated through independent component analysis (ICA) and singular value decomposition and subsequently subtracted from the signals. Finally, a second ICA was employed to enhance the fetal ECG signal.

The fetal ECG signals of all datasets were resampled to 500 Hz to give a common reference. Then they were divided in segments of 1920 samples and finally normalized to have zero mean and unity standard deviation. The normalization was performed in each segment separately.

2.5. Network training

In order to train the weights of the convolutional and the transposed convolutional kernels a loss function was minimized. A standard loss function used in denoising optimization problems is the mean squared error (MSE):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left(X^{(i)} - \hat{X}^{(i)}\right)^2 \tag{1}$$

where N is the number of training data in a batch, $\hat{X}^{(i)}$ is the ith denoised fetal ECG and $X^{(i)}$ is the clean version of $\hat{X}^{(i)}$, used as ground truth. We found that even when we normalized our data with zero mean and a standard deviation of one there were still some scale differences among the data, since we combined different datasets. To avoid favoring the data with the higher scale, instead of MSE we adopted as the loss function the normalized mean squared error (NMSE):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(X^{(i)} - \hat{X}^{(i)}\right)^2}{\bar{X}^{(i)2}} \tag{2}$$

where we normalized the MSE by the mean squared amplitude of the target signal $\bar{X}^{(i)2}$. We found that training with NMSE led to an improved network performance. The Adam algorithm (Kingma and Ba 2012) was selected as the optimization algorithm.

The simulated dataset was separated in two parts for testing and training. In the training set we added the signals generated with the modified fecgsyn toolbox based on VCGs 1–7, while the data based on VCGs 8 and 9 were assigned to the test set. The data simulated from adult ECG were also carefully separated into a training and test set. In addition, since the VCGs from the fecgsyn toolbox came from fitting a model on nine subjects from the PTB Diagnostic ECG Database, only these subjects were excluded to avoid any mixing of the training and test sets. Finally, the total number of training segments used in our experiments was 840 000 and there were 200 000 test segments. The network was trained on data with noise from −15 dB to 15 dB for 75 epochs until convergence was reached.

2.6. Performance measure

The performance of the network was assessed based on the improvement in the SNR achieved after the denoising compared with the SNR of the corrupted fetal ECG. The metric is computed as follows:

$$\begin{aligned} SNR_{imp}\,[dB] &= SNR_{post\text{-}denoising} - SNR_{pre\text{-}denoising} \\ &= 10\log_{10}\frac{\sum_{m=1}^{M}|Y[m]-X[m]|^2}{\sum_{m=1}^{M}\left|\hat{X}[m]-X[m]\right|^2} \end{aligned} \tag{3}$$

where $Y\,[m]$ denotes the corrupted fetal ECG signal and M is the length of the signals.

The proposed method was compared with two other ECG enhancement algorithms. The first one is beat-to-beat averaging, where we averaged 30 successive ECG complexes, similar to the STAN method (Hulsenboom *et al* 2019). The R peaks were detected before the averaging with the Pan–Tompkins algorithm (Pan and Tompkins 1985) in the clean signals. In this way, the averaging performance was not affected by the performance of the R-peak detector. The second method is a wavelet denoising method that decomposes the signal, thresholds the detail coefficients and reconstructs the signal to obtain its enhanced version. The symlet wavelet with six vanishing moments was selected as a mother wavelet because of its similarity to an actual ECG. A fixed threshold was used which was estimated by the minimax principle (Sardy 2000).

3. Results

3.1. Evaluation on simulated data

The denoising network was evaluated in the simulated test dataset. Figure 3 illustrates the performance of the developed fetal ECG denoising network compared with the 30-complex averaging and wavelet denoising methods. The input SNR refers to the SNR of the fetal ECG signals prior to denoising. As we can see in figure 3, the denoising network provides a considerable improvement in the SNR throughout the whole range of input

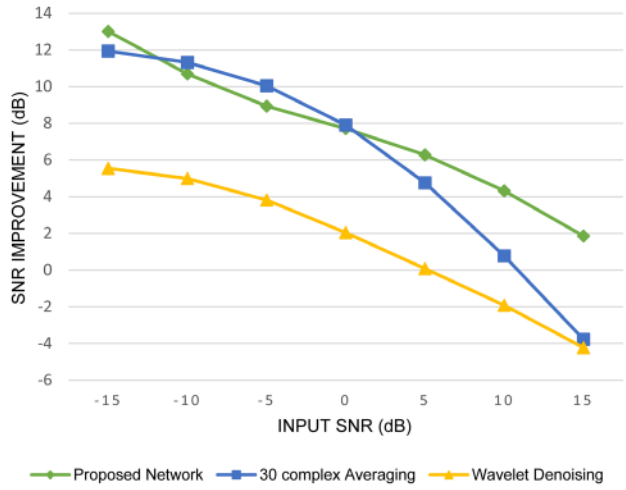


Figure 3. Performance of the proposed fetal ECG denoising network in comparison with other denoising methods in terms of improvement in the SNR of the denoised signal when compared with the noisy one.

SNR. In higher SNR ranges (more than 0 dB) our network outperforms the wavelet denoising and averaging techniques. This was expected since the averaging method does not preserve individual variations in the ECG complexes while our method is capable of doing so. Moreover, wavelet denoising distorts the signal amplitude whereas the denoising network does a better job in preserving it, especially in cases of low noise. For lower SNR values averaging performs slightly better than our method. The averaging method uses a larger signal segment for denoising (30 complexes) compared with the network (5–10 complexes). Apparently, an even larger effective denoising patch is needed for cases of heavy noise. On the other hand, unlike the averaging method, the proposed network does not make use of any prior information about the input signal. It should be mentioned here that the averaging method assumes that the R peaks are known. However, it is not guaranteed that the R peaks can be accurately estimated in very noisy signals. The strength of our method is that no prior processing of the signals, such as detecting the R peaks, is required and consequently errors in denoising originating from detection of the wrong peak are avoided.

Figure 4 depicts some typical results obtained in the test dataset for a wide range of SNR values. For relatively high SNR values (cases (a)–(c)) the network suppresses the noise to the extent that the ECG waves become distinguishable. Moreover, there is no or little distortion in the signal amplitude. For lower SNR values, around −5 dB, the QRS complexes are usually successfully recovered, as can be observed in figures 4(d) and (e). However, there are times when some P or T waves are created by the network where they do not actually exist. For example, in figure 4(d) the P waves in the denoised signal differ from the ground truth ECG in both location and polarity. In other instances the waves are difficult to recover or their amplitude is distorted (figure 4(e)). Finally, for very low SNRs (less than –10dB) even the QRS complexes appear to be challenging to recover (figure 4(f)). Thus, in low SNR ranges of the input signals even though the SNR of the output signal is significantly increased some morphological features might not always be reliably extracted.

Figure 5 illustrates the potential of our denoising method in a case with arrhythmia. The quality of the output signal has been enhanced to a great extent and nearly all the individual variations among the complexes are preserved. Unlike our algorithm, the averaging of successive complexes fails to maintain abrupt changes in signal morphology.

3.2. Performance on real fetal ECG signals

The fetal ECG denoising network was merely trained on simulated data. Therefore, it is interesting to examine how well it performs in terms of enhancing the quality of real fetal ECGs. Since ground truth real signals are not available, the performance of the network was tested by comparing the denoised signals of the Abdominal and Direct Fetal Electrocardiogram Database with the simultaneously recorded scalp ECG. It should be noted that the scalp ECG is a different ECG lead from the abdominal leads and that they should not be identical, even with perfect denoising. Nevertheless, the individual ECG segments should coincide between the abdominal leads and the scalp lead. Two intervals were computed and compared between the denoised signals and the scalp ECG in our analysis: the QT interval and the PR interval. The QT interval is defined as the interval from the onset of the QRS complex to the end of the T wave (offset of ventricular repolarization). The PR interval corresponds to the period that extends from the beginning of the P wave (onset of atrial depolarization) to the beginning of the QRS complex (onset of ventricular depolarization).

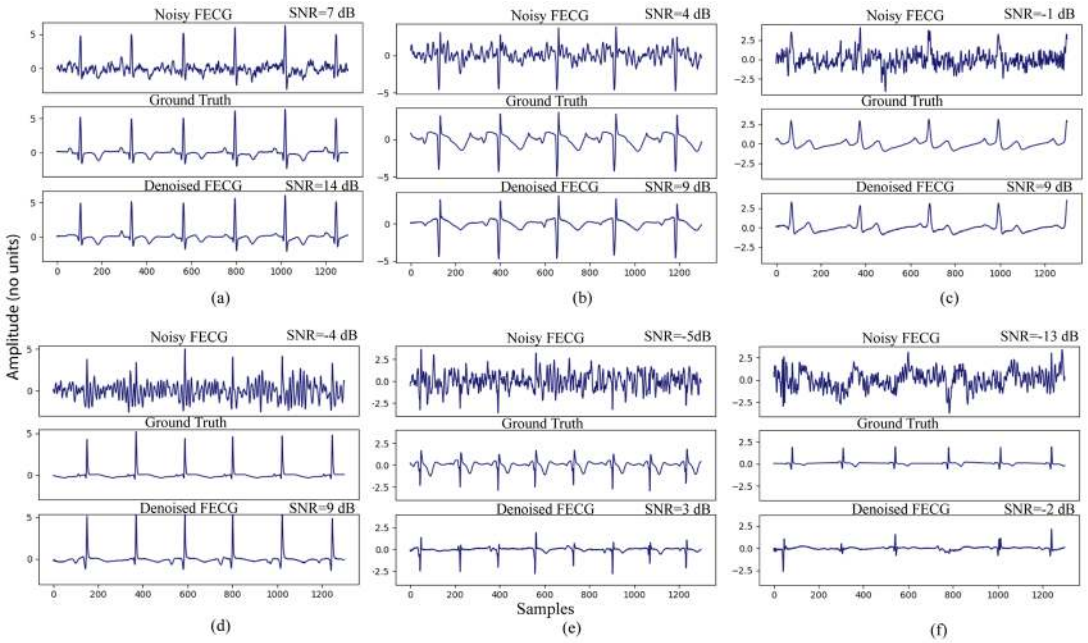


Figure 4. Denoising results for simulated fetal ECG signals in the test dataset with different amount of added noise. For relatively high input SNR the network achieves a high-quality result, while for lower SNR values some morphological features might be unreliable.

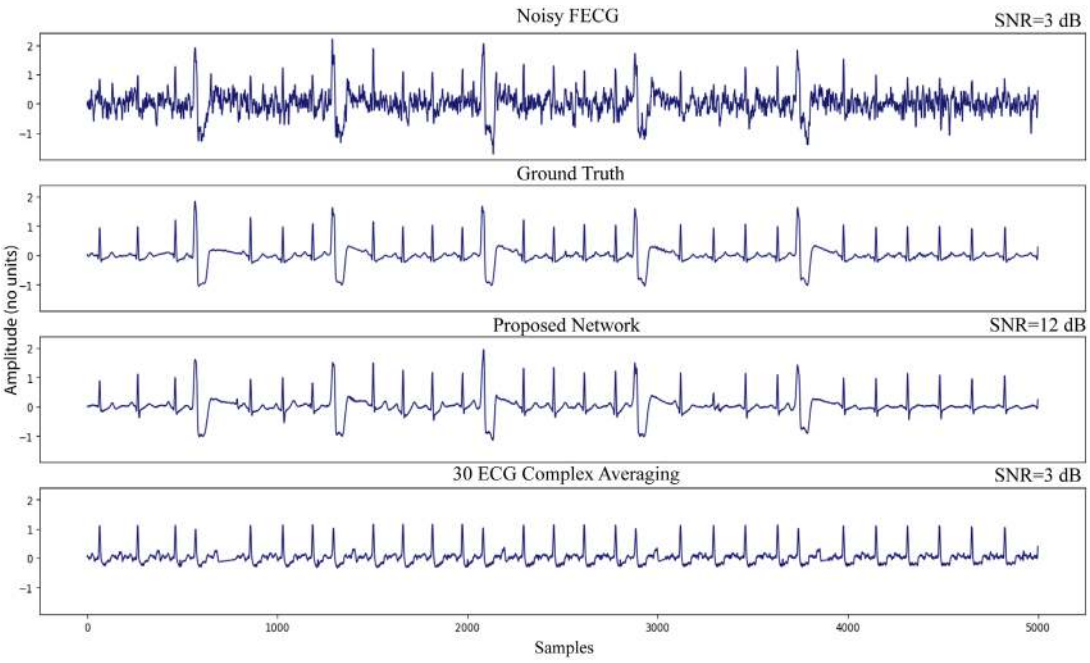


Figure 5. An example of denoising a simulated fetal ECG exhibiting arrhythmia with the proposed deep CNN network compared with the 30 ECG complex averaging method. Unlike averaging, the suggested network can preserve diagnostic morphological information of the signal.

The scalp measurements contain a considerable amount of noise, making it challenging to determine the intervals. To enable interval detection, quality enhancement of the scalp ECG was necessary. For this purpose, we performed averaging of seven successive ECG complexes. In order to have a fair comparison, the same procedure was applied to the denoised signals. Even after averaging of seven complexes there was substantial noise left in some segments of the scalp ECGs, therefore we had to exclude them from the comparison. However, by averaging more complexes we risked assessing the performance of the averaging method rather than our proposed denoising. Choosing to use seven complexes was a compromise between sufficient quality enhancement of the scalp ECGs and preservation of variation among consecutive complexes.

Table 1. Comparison of the PR and QT intervals between the denoised fetal ECG signals and the scalp ECG. All the values are expressed in ms.

Interval	RMSE	R^2	RMSE95	$R^2$95
PR	9.9	0.77	7.4	0.86
QT	14	0.81	11	0.87

The end of the T wave is difficult to define (Rautaharju *et al* 2009), and to calculate it we adopted the tangent technique. According to this technique, a tangent line is determined down the steepest slope of the terminal limb of the T wave. Afterwards, the end of the T wave is defined by the intersection of this line with the baseline. The beginning of the P wave was determined in a similar way. The intervals were calculated first in an automated way and then checked manually by an expert to correct for errors. From each recording, only the channel with the smallest amount of noise was analyzed. Since the recordings are 5 min long this corresponds to approximately 620 intervals per recording (about 3100 in total). The signal parts where the amount of noise was too high or the intervals were not recognizable were excluded from the comparison (for both the scalp and denoised signals). After this exclusion, 2000 intervals were compared.

The root mean square error (RMSE) and the squared Pearson correlation coefficient (R^2) were computed to measure how well the intervals match between the scalp and the denoised fetal ECG signals. Moreover, we evaluated the RMSE95 and $R^2$95 where the extreme 5% values were excluded. This was done to ensure that the outliers did not bias the estimation of the metrics. Table 1 displays the calculated values for the two intervals. The RMSE for the PR interval was estimated to be 9.9 ms, while after the exclusion of the extreme values it was 7.4 ms. For the QT interval RMSE is equal to 14 ms and RMSE95 11 ms. We should note here that in the scalp measurements the average value for the PR interval was found to be 120 ms and for the QT interval 270 ms. For the PR interval the squared correlation coefficient is 0.77 while for the QT it is 0.81. The removal of the extreme values raised the value to 0.86 for the PR interval and 0.87 for the QT interval. We also provide the correlation and the Bland–Altman plots for the estimated intervals in figure 6. The Bland–Altman plot shows that the PR interval was measured to be an average of 6 ms shorter in the denoised signals compared with the scalp electrode. The QT interval was computed to be 6.1 ms longer in the non-invasive fetal ECG signals. Moreover, we observe that the limits of agreement are not very wide, especially for the QT interval, but there are measurements that fall outside them. This means that denoising does not provide consistently good results but has some outliers.

Finally, a visual result of the denoising is provided in figure 7 for channel 2 of the recording 'r04' of the Abdominal and Direct Fetal Electrocardiogram Database. For better visualization the scalp ECG was high-pass filtered to remove the baseline wander. The denoised ECG signal is free from noise and the individual waves correspond relatively well to those in the scalp ECG.

4. Discussion

This paper presents a deep CNN network for denoising fetal ECG signals. Fetal ECG signals, obtained from non-invasive recordings, contain a substantial amount of noise even after the maternal ECG has been suppressed. This renders it difficult for clinicians to examine and interpret the morphology of the ECG signals. Usually, the QRS complex can be detected without further processing, due to the high amplitude of the R peak. However, the smaller waves, like the P and T waves, can often not be readily distinguished. The proposed CNN network was developed to post-process the extracted fetal ECG signals in order to further enhance their quality. The network was trained merely on simulated data and its performance was examined both on synthetic and real signals.

The network was found to be efficient in reducing residual noise in the synthetic ECG dataset. When the SNR of the signals was relatively high the T and P waves were mostly reconstructed with no or little distortion. In most cases the waves in the noisy input signals were impossible or too difficult to detect. However, in cases with low SNR, even if a significant improvement in the SNR was achieved some morphological features were distorted, absent or 'fake'. This means that the network cannot reliably reconstruct the ECG morphology when the signals are severely corrupted by noise. It is possible that signal segments longer than 4 s are needed to capture more information about the underlying signal structure. This finding suggests that if the developed network is intended to be used in practice then prior quality assessment of the input signal is necessary as a reliability measure of the result. If the quality of the signals is assessed as low, then the network should not be used for denoising. Alternatively, longer signal segments can possibly be considered and more layers should be added to the network to achieve a larger effective denoising patch at the expense of higher computational intensity and the possible need for even more data. Moreover, in our analysis we have only experimented with single-channel ECGs. However, modifying our network to handle multichannel fetal ECGs would be straightforward by employing 2D convolutions. Allowing the convolutions to extend in both space and time might lead to more accurate results.

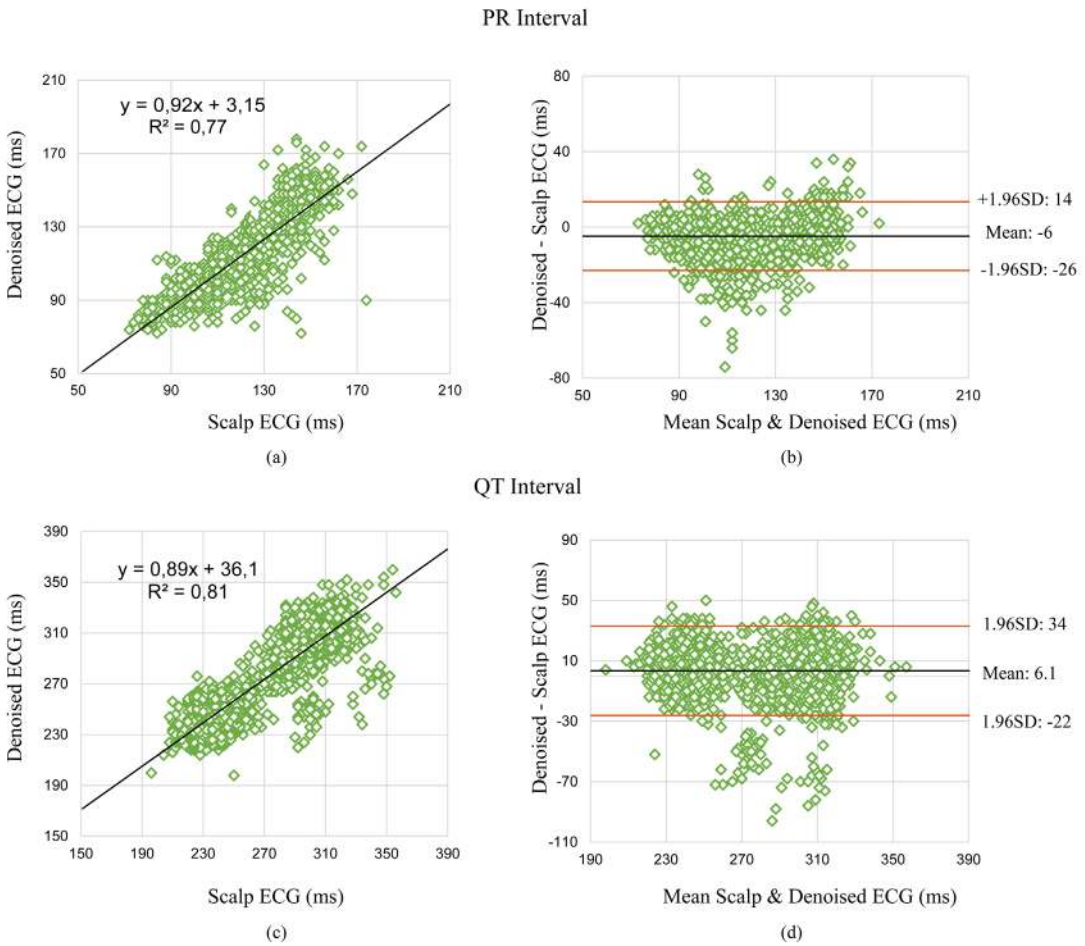


Figure 6. Interval comparison between the denoised non-invasive fetal ECG signals and the simultaneously recorded scalp ECG: (a), (c) correlation plots and (b), (d) Bland–Altman plots for the PR and QT intervals, respectively. In (a) and (c) a line fit is given by $y =$ intercept + gradient x and R^2 is the corresponding squared Pearson correlation coefficient.

In a real dataset we measured how well the PR and QT intervals of the denoised signals correspond to those measured in the scalp ECG. The intervals were extracted in the signals created as a running average of seven heartbeats. There is a limitation in this comparison, since there is some further denoising performed because of the averaging. However, the need to remove the noise in the scalp, to obtain reliable intervals, outweighs the mentioned limitation. We found a RMSE of 9.9 ms for the PR interval and 14 ms for the QT interval. These values show high similarity between the two measurements, especially if we consider errors and high ambiguity in the computation of the intervals. The value for the QT interval is similar that found by Behar *et al* (2016) (13.6 ms) where annotations of several cardiologists were combined and compared with the scalp ECG. However, compared with Behar *et al* (2016), we used fewer subjects in our comparison. Furthermore, the running average computation in Behar *et al* (2016) was different from ours since they selected only those ECG complexes with high similarity for the averaging in 1 min segments. They found that the QT interval can be reliably extracted only if computed as a running average of several heartbeats. In our case the averaging is less necessary since most of the noise is removed by the denoising network.

It bears mentioning, however, that validation in a more extensive real dataset is needed to confirm our findings. Moreover, the signals in our study were recorded between 38 and 41 weeks of gestation. This is because it is impossible to have scalp measurements before birth. However, it would be interesting to assess the accuracy of the method in earlier gestational ages as well.

Despite the limitations in the validation of our method, this is the first study to show the potential of deep CNNs for efficient removal of noise from non-invasive fetal ECG signals. The principal advantage of this method over the widely used running average method is that no prior R-peak detection is necessary and that individual variations in pulse shape and beat-to-beat interval are retained. This is especially beneficial in arrhythmia cases. Nowadays, the averaging performed in the fetal ECG precludes its use in real-time arrhythmia analysis, and hence fetal arrhythmia can only be assessed through echocardiography. This work brings us a step closer to broadening our understanding of the mechanisms of fetal arrhythmia by examining the fetal ECG. Moreover, the quality of the denoised signals is high enough to allow us to measure the exact timing of different morphological features

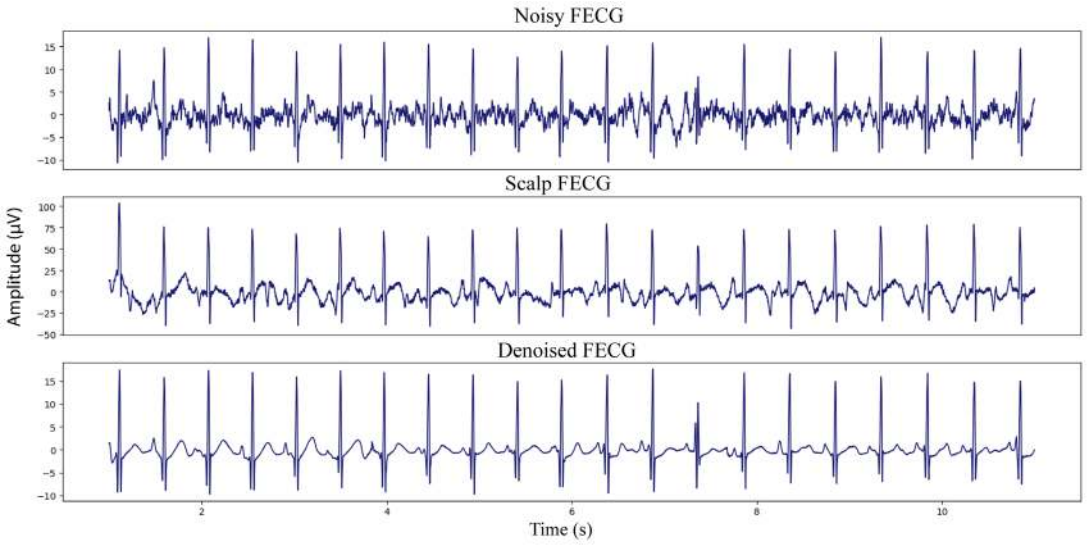


Figure 7. The denoising result of the proposed network for channel 2 of recording 'r04' of the Abdominal and Direct Fetal Electrocardiogram Database. The enhanced signal is clean and the morphological signal characteristics are similar to those of the scalp ECG.

of the ECG signal by clinicians. Besides, it can facilitate and advance research towards automated detection of fetal ECG intervals and segments. Extracting morphological features from the ECG signal allows for estimation of the well-being of the fetus. Metabolic acidosis was found to be associated with variations in QT length (Oudijk *et al* 2004). Several studies have also demonstrated a physiologically negative correlation between the PR and RR intervals which becomes positive with evolving acidosis (Murray 1986, van Wijngaarden *et al* 1996). However, the role of ECG intervals in fetal monitoring is yet to be established, but for this to be achieved the technological limitations related to non-invasive ECG need to be overcome.

5. Conclusion

In this paper we propose a deep encoder–decoder framework for non-invasive fetal ECG signal denoising. Convolutions and transposed convolutions are combined to remove the noise by extracting primary signal content and recovering details. Experimental results in synthetic signals showed that the network is able to achieve a substantial improvement in quality of the noisy signals. However, when the signals are heavily corrupted by noise some morphological features are unreliable, stressing the need for a reliability measure of the network's output. Experiments on real signals demonstrated high correlation of the PR and QT intervals in the denoised signals when compared with the scalp ECG. The principal advantage of the method is that individual variations among different pulses can be preserved and that, unlike most other fetal ECG denoising methods, this method does not require knowledge of R-peak locations.

Conflict of interest

RV has shares in Nemo Healthcare BV, The Netherlands.

ORCID iDs

Eleni Fotiadou https://orcid.org/0000-0003-3877-8961
Jürgen Hesser https://orcid.org/0000-0002-4001-1164

References

Adam D and Shavit D 1990 Complete foetal ECG morphology recording by synchronized adaptive filtration *Med. Biol. Eng. Comput.* **28** 287–92
Andreotti F *et al* 2013 Maternal signal estimation by Kalman filtering and template adaptation for fetal heart rate extraction *Computing in Cardiology 2013 (Zaragosa)* pp 193–6
Andreotti F *et al* 2016 An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms *Physiol. Meas.* **37** 627–48
Antczak K 2018 Deep recurrent neural networks for ECG signal denoising arXiv 2018 (arXiv:1807.11551)
Behar J *et al* 2013 Non-invasive FECG extraction from a set of abdominal sensors *Computing in Cardiology 2013 (Zaragosa)* pp 297–300

Behar J *et al* 2016 Evaluation of the fetal QT interval using noninvasive fetal ECG technology *Physiol. Meas.* **37** 1392–403
Bousseljot R *et al* 1995 Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet *Biomed. Eng.-Biomed. Te.* **40** 317–8
Burger H C *et al* 2012 Image denoising: can plain neural networks compete with BM3D? *IEEE Conf. Comput. Vision and Pattern Recognition (Providence, RI)* pp 2392–9
Camargo-Olivares J L *et al* 2011 The maternal abdominal ECG as input to MICA in the fetal ECG extraction problem *IEEE Signal Process. Lett.* **18** 161–4
Cerutti S *et al* 1986 Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings *J. Perinat. Med.* **14** 445–52
Clifford G D *et al* 2014 Non-invasive fetal ECG analysis *Physiol. Meas.* **35** 1521–36
Fotiadou E *et al* 2018 Enhancement of low-quality fetal electrocardiogram based on time-sequenced adaptive filtering *Med. Biol. Eng. Comput.* **56** 2313–23
Goldberger A L *et al* 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20
He K *et al* 2016 Deep residual learning for image recognition *Proc. IEEE Conf. Comp. Vis. Patt. Recogn. (Las Vegas)* pp 770–8
Hulsenboom A D J *et al* 2019 Relative versus absolute rises in T/QRS ratio by ST analysis of fetal electrocardiograms in labour: a case-control pilot study *PLoS One* **14** e0214357
Isin A and Ozdalili S 2017 Cardiac arrhythmia detection using deep learning *Procedia Comput. Sci.* **120** 268–75
Jain V and Seung S 2008 Natural image denoising with convolutional networks *Proc. Advances in Neural Inf. Process. Syst. (Vancouver)* pp 769–76
Jezewski J *et al* 2006 Comparison of Doppler ultrasound and direct electrocardiography acquisition techniques for quantification of fetal heart rate variability *IEEE Trans. Biomed. Eng.* **53** 855–64
Jezewski J *et al* 2012 Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram *Biomed. Eng.* **57** 383–94
Kingma D and Ba J 2012 Adam: a method for stochastic optimization *Int. Conf. for Learning Representations (San Diego)* pp 1–13
Laguna P *et al* 1997 A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG *Computers in Cardiology* pp 673–6
Lee J S *et al* 2018 Fetal QRS detection based on convolutional neural networks in noninvasive fetal electrocardiogram *4th Int. Conf. on Frontiers of Signal Processing (Poitiers)* pp 75–8
Lindecrantz K 1983 *Processing of the Fetal ECG; an Implementation of a Dedicated Real Time Microprocessor System* (Gothenburg: Chalmers University of Technology)
Lu X *et al* 2013 *Speech Enhancement Based on Deep Denoising Autoencoder* (Lyon: INTERSPEECH)
Maas A *et al* 2012 *Recurrent Neural Networks for Noise Reduction in Robust ASR* (Portland, OR: INTERSPEECH)
Mao X *et al* 2016 Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections *NIPS (Barcelona)* pp 2802–10
Martín-Clemente R *et al* 2002 Fast technique for noninvasive fetal ECG extraction *IEEE Trans. Biomed. Eng.* **58** 227–30
Murray H G 1986 The fetal electrocardiogram: current clinical developments in Nottingham *J. Perinat. Med.* **14** 399–404
Niknazar M *et al* 2013 Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings *IEEE Trans. Biomed. Eng.* **60** 1345–52
Oudijk M A *et al* 2004 The effects of intrapartum hypoxia on the fetal *QT* interval *Br. J. Obstet. Gynaecol.* **111** 656–60
Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **BME-32** 230–6
Podziemski P and Gieraltowski J 2013 Fetal heart rate discovery: algorithm for detection of fetal heart rate from noisy, noninvasive fetal ECG recordings *Computing in Cardiology 2013 (Zaragosa)* pp 333–6
Rautaharju P M *et al* 2009 AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram *J. Am. Coll. Cardiol.* **53** 982–91
Sameni R 2008 Extraction of fetal cardiac signals from an array of maternal abdominal recordings *PhD Thesis* Sharif University of Technology—Institut National Polytechnique de Grenoble
Sardy S 2000 Minimax threshold for denoising complex signals with Waveshrink *IEEE Trans. Signal Process.* **48** 1023–8
Signorini M G *et al* 2003 Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings *IEEE Trans. Biomed. Eng.* **50** 365–74
Silva I *et al* 2013 Noninvasive Fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013 *Comput. Cardiol.* **40** 149–52
Srivastava R K *et al* 2015 Training very deep networks *Adv. Neural. Inf. Process. Syst. (Montreal)* pp 2377–85
Ungureanu M *et al* 2007 Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique *Biomed. Technol.* **52** 56–60
van Laar J O *et al* 2014 Fetal heart rate variability during pregnancy, obtained from non-invasive electrocardiogram recordings *Acta Obstet. Gynecol. Scand.* **93** 93–101
van Wijngaarden W J *et al* 1996 Changes in the PR interval–fetal heart rate relationship of the electrocardiogram during fetal compromise in chronically instrumented sheep *Am. J. Obstet. Gynecol.* **175** 548–54
Varanini M *et al* 2013 A multi-step approach for non-invasive fetal ECG analysis *Computing in Cardiology 2013 (Zaragosa)* pp 281–4
Verdurmen K M J *et al* 2016 Normal ranges for fetal electrocardiogram values for the healthy fetus of 18–24 weeks of gestation: a prospective cohort study *BMC Pregnancy Childbirth* **16** 227
Vullings R *et al* 2009 Dynamic segmentation and linear prediction for maternal ECG removal in antenatal abdominal recordings *Physiol. Meas.* **30** 291–307
Widrow B *et al* 1975 Adaptive noise cancelling: principles and applications *Proc. IEEE* **63** 1692–716
Xiong P *et al* 2016 ECG signal enhancement based on improved denoising auto-encoder *Eng. Appl. Artif. Intell.* **52** 194–202
Ye Y *et al* 2009 An efficient semi-blind source extraction algorithm and its applications to biomedical signal extraction *Sci. China* F **52** 1863–74
Zhang H *et al* 2009 Semi-blind source extraction algorithm for fetal electrocardiogram based on generalized autocorrelations and reference signals *J. Comput. Appl. Math.* **223** 409–20
Zhang K *et al* 2017 Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising *Trans. Image Process.* **26** 3142–55
Zhong W *et al* 2018 A deep learning approach for fetal QRS complex *Physiol. Meas.* **39** 045004