

Fetal Electrocardiogram Extraction Using Dual-Path Source Separation of Single-Channel Non-Invasive Abdominal Recordings

Arash Shokouhmand , Student Member, IEEE, and Negar Tavassolian , Senior Member, IEEE

Abstract—Objective: The development of a method for non-invasive monitoring of fetal electrocardiogram (FECG) signals from single-channel abdominal recordings. **Methods:** The dual-path source separation (DPSS) architecture is introduced for the simultaneous separation of fetal and maternal ECG signals from abdominal ECG recordings. DPSS initially denoises abdominal ECG (AECG) recordings using a generative dual-path long short-term memory (DP-LSTM) network. An inception module along with a series of DP-LSTM blocks is then employed to extract the mask-ing maps associated with fetal and maternal components. Finally, these masking maps are weighted by the AECG recording to separate maternal and fetal ECG signals. The performance of this network is evaluated on 10 pregnancies from the fetal ECG synthetic database (FECGSYNDB), 22 cases of labor and pregnancy from the abdominal and direct fetal ECG database (ADFECGDB), and 69 pregnancies from set A of non-invasive FECG challenge (NIFECGC) datasets. **Results:** F1-scores of 99.03%, 97.7%, and 95.3% are reported for the detection of fetal QRS complexes in FECGSYNDB, ADFECGDB, and NIFECGC respectively. DPSS technique is also evaluated in terms of separability of fetal and maternal clusters. According to the clustering-based analyses, the average purity index of 0.9750, Jaccard index of 0.9705, and Davies-Bouldin index of 0.7429 demonstrate the high source separation capability of DPSS. **Conclusion:** The achieved performance suggests that DPSS enables accurate single-channel FECG extraction, and can replace state-of-the-art source separation techniques for this purpose. **Significance:** This study signifies a fundamental step towards non-invasive fetal ECG monitoring systems, which favors at-home prenatal care.

Index Terms—Fetal ECG monitoring, neural networks, non-invasive recordings, single-channel monitoring, source separation.

Manuscript received 9 May 2022; revised 18 June 2022; accepted 6 July 2022. Date of publication 11 July 2022; date of current version 26 December 2022. This work was supported by the National Science Foundation (NSF) under Award 1855394. (Corresponding author: Negar Tavassolian.)

Arash Shokouhmand is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, USA.

Negar Tavassolian is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: negar.tavassolian@stevens.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBME.2022.3189617>, provided by the authors.

Digital Object Identifier 10.1109/TBME.2022.3189617

I. INTRODUCTION

STILLBIRTH, defined as the death of a fetus after 24 weeks of gestational age [1], affects 2.5 million babies globally every year [2]. Stillbirth is often accompanied by a decrease in fetal heart rate (FHR) [1], provoking the need for proactive fetal monitoring to reduce fetal mortality. Current fetal heart rate monitoring predominantly relies on clinical cardiotocography (CTG) systems which are equipped with an ultrasound transducer [3]. Alternatively, handheld 1D Doppler ultrasound devices are used to monitor fetal heart rate during pregnancy [4]. The ultrasound probe is placed on the abdominal region, which emits 1–3 MHz sound waves towards the fetus. The fetal heart rate is estimated based on the Doppler frequency shift in the back-scattered signal due to the moving tissues of the fetal cardiac muscles.

Prenatal complications such as stillbirth may arise in between clinical visits when timely medical interventions could alleviate the risk of adverse outcomes. This requires continuous tracking of the fetal heart rate which serves expectant mothers to be informed of the wellbeing of the fetus. Although ultrasound is the most common technology for measuring fetal heart rate, it is technically limited by frequent periods of signal loss in the cases of premature deliveries and high body-mass-index (BMI) mothers [5], [6]. Therefore, it is not suitable for ubiquitous fetal monitoring. Furthermore, the food and drug administration (FDA) recommends against the use of home-based ultrasound technology due to its potential harm to the fetus and eventually the mother [7].

In the past few years, the advent of wearable sensors has allowed for the dissemination of wearable health monitoring devices [8]. In parallel, a variety of technologies including electrocardiogram (ECG) [9], phonocardiogram (PCG) [10], seismocardiogram (SCG) [11], and gyro-cardiogram (GCG) [12] have been proposed for FHR monitoring. Non-invasive fetal ECG (NIFECG) has recently shown promise for FHR monitoring [13]. This technology incorporates ECG electrodes placed on the mother's abdomen to acquire abdominal ECG signals [14]. However, these recordings consist of several components including maternal ECG, fetal ECG, uterine contractions, fetal movements, and noise components which complicate FHR estimation [15].

Blind source separation (BSS) techniques such as principal component analysis (PCA) [16], periodic component analysis (π CA) [17], and independent component analysis (ICA) [18]

were used to separate fetal ECG components from abdominal recordings. BSS methods aim to decompose abdominal ECG signals into statistically independent sources including FECG, MECG, and noise. Despite promising results, BSS methods require the number of ECG sensors to be equal to or greater than the number of independent sources [19].

Adaptive filters signify another means of fetal ECG extraction. These filters leverage a transfer function with variable coefficients trained to remove maternal ECG from abdominal recordings. The least mean squares (LMS) filter and recursive least squares (RLS) filter represent two types of adaptive filtering applicable to fetal ECG extraction [20]–[23]. Authors in [24] adopted an echo state neural network (ESN) to project chest-recorded MECG signals onto a set of non-orthogonal basis functions, where an RLS algorithm was used to update the weights of the network and estimate the fetal ECG signal with an F1-score of 90.2%. The LMS and RLS filters require either a reference maternal ECG channel that morphologically resembles the abdominal recordings, or several linearly independent channels to reconstruct abdominal recordings. The need for reference sensors limits the use of adaptive filters [25]. Alternatively, Kalman filtering approaches require only an arbitrary maternal ECG reference for fetal ECG enhancement. The performance of this method, however, is limited when fetal and maternal QRS peaks coincide [26]. This drawback was addressed by the extended-state Kalman filters which enable separating coincided maternal and fetal peaks at the expense of high computational complexities [27].

Template subtraction (TS) strategies aim to enhance fetal ECG by mitigating the impact of maternal ECG cycles in abdominal recordings. To this end, a maternal ECG template is obtained by averaging the preceding maternal ECG cycles, which is often followed by one or more of the following techniques for fetal ECG extraction: subspace reconstruction by singular value decomposition (SVD) [28], ICA [29], or PCA [16], temporal subtraction of the maternal template [30], and filtering [31]. TS methods can be applied to both single-channel and multiple-channel recordings. In order to use template subtraction methods however, a few assumptions including independence of fetal and maternal components and the consistency of heartbeat cycles in fetal and maternal ECG signals, are made which do not necessarily hold true for abdominal recordings [32]. Having analyzed the traditional methods, we revisit the estimation of fetal ECG and consider it as an under-determined problem aiming to extract fetal ECG signals from single-channel abdominal recordings.

The structure of the paper is organized as follows: In Section II, previous studies on single-channel fetal ECG extraction are reviewed. Section III details the proposed methodology for single-channel fetal ECG extraction. In Section IV, experimental results on real and synthetic datasets are discussed. The paper is concluded in Section VI, where future directions are also discussed.

II. SINGLE-CHANNEL STATE-OF-THE-ART

Adaptive filters have recently been proposed for fetal ECG extraction from single-channel recordings [33]–[39]. The authors in [39] introduced a convolution layer which aims to find

the maternal ECG signal as a reference for an RLS adaptive filter. As such, fetal ECG QRS peaks were obtained with an F1-score of 96.0% by subtracting MECG components from abdominal recordings.

Increasing the number of channels using signal decomposition methods is another strategy proposed for fetal ECG extraction. The authors in [40]–[42] use empirical mode decomposition (EMD) or SVD to decompose the signal or its spectrogram into multiple channels, which is then followed by independent component analysis (ICA) or non-negative matrix factorization (NMF) for fetal ECG separation. These methods are limited to the assumption that the generated channels are statistically independent. Furthermore, some works use decomposition techniques in conjunction with peak detection and coherent averaging to cancel maternal ECG [43], [44]. Recently, [45] proposed to factorize the STFT of the AECG signals using NMF, and extract FECG and MECG by setting threshold levels on the activation matrix for maternal and fetal components. As reported, average F1-scores of 94.80% and 84.0% were achieved for the ADFECGDB and NIFECGC datasets, respectively. In [46], the Shannon energy of abdominal time-frequency representation is used to localize maternal ECG peaks for template subtraction, achieving F1-scores of 98.67% and 99.27% on NIFECGC and ADFECGDB, respectively. The performance of threshold-based methods, however, can be affected if fetal and maternal QRS-peaks coincide, or the abdominal segment length is not long enough to accurately estimate the maternal ECG template.

In recent years, the use of machine learning techniques has attracted extensive attention for fetal ECG monitoring [47]–[50]. Authors in [47] leveraged a clustering method to classify singularities from abdominal ECG recordings into noise, fetal QRS, and maternal QRS peaks. This method reported F1-scores of 98.04% and 98.63% for ADFECGDB and NIFECGC datasets, respectively. Furthermore, a residual convolutional encoder-decoder network (RCED-Net) was introduced in [48] to separate fetal ECG components from abdominal recordings, where F1-scores of 94.10% and 93.62% were achieved for ADFECGDB and NIFECGC, respectively. Recently in [49], a domain adaptation method was proposed to map the abdominal ECG recordings to their FECG counterparts using a cycle generative adversarial network (CycleGAN). The authors reported an F1-score of 94.7% for the NIFECGC database. A combination of STFT and generative adversarial networks (GANs) was employed in [50] to directly map abdominal recordings to their fetal ECG counterparts. The performance of this method is limited to F1-scores of 93.02% on NIFECGC and 90.05% on ADFECGDB.

In this study, we introduce a single-channel source separation approach based on the dual-path modeling of abdominal ECG signals. This method is coined dual-path source separation (DPSS) architecture, and leverages long short-term memory (LSTM) cells to discriminate the components of FECG and MECG from each other. In our work, the dual-path recurrent neural network (DPRNN), initially introduced in speech processing [52], is utilized to model temporal information of the abdominal ECG signal to denoise and separate fetal ECG components from those of maternal ECG. The effectiveness of the

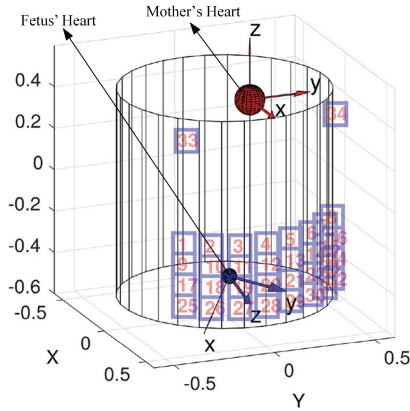


Fig. 1. The distribution of ECG sensors on the abdominal region in FECGSYNDB dataset [51].

proposed framework is assessed on real and synthetic datasets as explained in the following section.

III. METHODOLOGY

A. Experimental Setup & Datasets

Our experimental setup incorporates non-invasive abdominal ECG recordings along with corresponding reference annotation of QRS complexes. To develop our experimental setup, three publicly available datasets including the fetal ECG synthetic database [51], [53], the abdominal and direct fetal ECG database [54]–[56], and the non-invasive fetal ECG challenge database [57], [58] from PhysioNet [59] were used, as described in the following sub-sections.

1) Fetal ECG Synthetic Database (FECGSYNDB): The fetal ECG synthetic database (FECGSYNDB) includes 32-channel artificial non-invasive abdominal ECG (AECG) recordings of 10 subjects, amounting to 145.8 hours of data. Along with AECG, the ground-truth maternal ECG (MECG) and fetal ECG (FECG) signals are also provided separately. A distinctive characteristic of FECGSYNDB is the inclusion of non-stationary events including fetal movements, uterine contractions, heart rate acceleration/deceleration, and ectopic heartbeats. These non-stationary events ensure simulating the dynamics of realistic pregnancies. Each recording consists of 5 minutes of data sampled at 250 Hz. In this study, we used the data of all 32 channels distributed over the abdominal region as shown in Fig. 1. In order to simulate realistic scenarios, pregnancy-specific noise components corresponding to fetal movement, uterine contractions, thermal interference, and maternal heartbeat from [51] are mixed with the fetal ECG signals to generate signal-to-noise ratios (SNR) within the range of $(-9, +2)$ dB. Fetal and maternal acceleration/deceleration scenarios are also incorporated to simulate abnormal scenarios.

2) Abdominal and Direct Fetal ECG Database (ADFECGDB): This dataset comprises 4-channel abdominal ECG recordings sampled at 500 Hz from 10 pregnancy and 12 labor cases between the 38 and 42 weeks of gestation at the Medical University of Silesia. Each labor signal consists of 5

minutes of ECG data acquired from maternal abdomen around the umbilicus (AECG). Additionally, a direct fetal electrocardiogram recorded from the scalp of the fetus (FECG) is provided as the reference signal. Each pregnancy recording includes twenty minutes of abdominal ECG in which fetal QRS peaks were annotated by an automated fetal QRS detector. Annotations of fetal QRS complexes for both labor and pregnancy datasets were validated by medical experts and provided in the dataset. While the whole dataset is represented by ADFECGDB, labor and pregnancy datasets are shown by ADFECGDB(L) and ADFECGDB(P) in this work, respectively.

3) Non-Invasive Fetal ECG Challenge (NIFECGC): Set A of the non-invasive FECG database from the PhysioNet/computing challenge is the largest publicly available dataset for non-invasive fetal monitoring. This collection consists of 75 4-channel abdominal ECG recordings sampled at 1 kHz with a signal duration of 1 minute. A team of expert clinicians have manually annotated the QRS complexes on the acquired signals. A few recordings (a33, a38, a47, a52, a54, a71, and a74) are excluded from the dataset due to inaccurate QRS annotations as suggested by [9] and [60].

In this study, only the channel with the most evident fetal ECG components was selected per subject in the NIFECGC and ADFECGDB datasets, as summarized in Tables I and II of the supplementary material respectively. In single-channel fetal ECG extraction, a signal quality index (SQI) such as that calculated by the naive Bayes classifier in [61] can be computed for the extracted fetal ECG signal to confirm the optimality of the ECG lead position. This mechanism also assists with adjusting the ECG electrodes on the abdominal region in case the signal quality degrades due to fetal movement.

B. Data Preparation

For the sake of uniformity, the NIFECGC and FECGSYNDB datasets were initially resampled to 500 Hz using fast Fourier transform (FFT). Abdominal recordings are often accompanied by baseline wander and power-line noise which should be removed [13]. To remove baseline wander, abdominal ECG signals were high-pass filtered using a zero-phase Butterworth filter with a cut-off frequency of 2 Hz. To attenuate power-line interference, a second-order infinite impulse response (IIR) notch filter centered at 50 Hz or 60 Hz with a quality factor of 25 was applied to the filtered AECG signals in forward and backward directions. Finally, each abdominal signal was segmented into 4-second overlapping windows with 90% overlap between consecutive windows. As such, each segment represents 2000 samples (4 seconds at 500 Hz) of data processed for fetal ECG separation. The resulting segments are then normalized into a standard distribution (zero-mean and unit variance) to ensure a stable convergence when training the network.

C. Fetal ECG Extraction

In this section, we describe the proposed approach based on the network illustrated in Fig. 2. This network consists of two main blocks, namely denoising and source separation blocks.

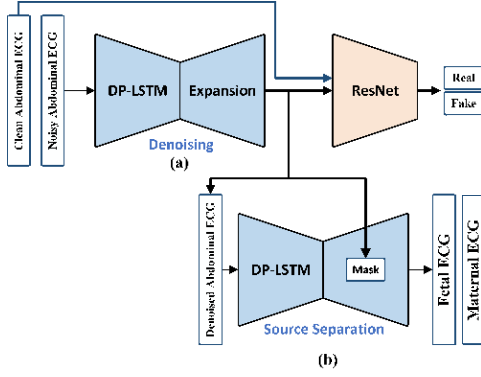


Fig. 2. Dual-path source separation (DPSS) framework including (a) a generative network for denoising abdominal ECG, and (b) a dual-path LSTM (DP-LSTM) network for the separation of fetal and maternal ECG signals.

Initially, the abdominal ECG signal is denoised using a generative network as shown in Fig. 2(a). The denoised abdominal ECG then undergoes a masking-based source separation network as shown in Fig. 2(b). This network extracts the fetal ECG signal by masking maternal components and the maternal ECG signal by masking fetal components. As the dual-path long short-term memory mechanism is used in many parts of the architecture, we first describe this module, and then explain the denoising and source separation networks.

1) Dual-Path LSTM Architecture: Long short-term memory (LSTM) networks have been useful for learning sequential data such as speech signals, text, and time series [62]. However, they fail to model long sequences due to vanishing gradients [63]. To address this issue in this work, the dual-path LSTM (DP-LSTM) network is proposed for modeling abdominal ECG signals as shown in Fig. 3. The intuition behind the DP-LSTM is modeling shorter segments (local modeling), rather than the whole signal, and then aggregating the information across all segments (global modeling). For this purpose, a sequential input $I_l \in \mathbb{R}^{M \times T}$ with M and T representing the number of input channels and timesteps respectively, is segmented into K segments of length L ($T = K \times L$). As such, a segmented feature map $S_l \in \mathbb{R}^{M \times K \times L}$ is obtained. Fig. 3(a) illustrates the local modeling path. In this path, a bi-directional LSTM (Bi-LSTM) module is employed to model each sequence forward and backward to capture past and future information, respectively. The output of this layer represents a matrix with a different number of channels, i.e., $B_l \in \mathbb{R}^{H \times K \times L}$. This difference occurs as the input and output vectors in an LSTM block are not necessarily of the same dimensions. To restore the original dimensions ($M \times K \times L$), a fully-connected (FC) layer is then applied to B_l , which results in $F_l \in \mathbb{R}^{M \times K \times L}$. In order to accelerate the learning process and enhance the generalizability of the model [64], a normalization layer is used as below:

$$N_l = \frac{F_l - E\{F_l\}}{\sqrt{\text{Var}\{F_l\}}} \gamma + \beta, \quad (1)$$

where $N_l \in \mathbb{R}^{M \times K \times L}$, $\gamma \in \mathbb{R}^{M \times 1}$, $\beta \in \mathbb{R}^{M \times 1}$, and represent the normalized matrix, normalization scaler, normalization bias,

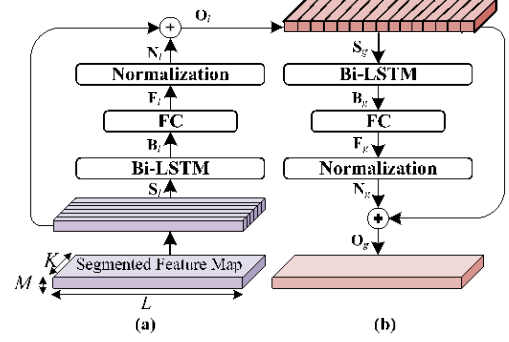


Fig. 3. Dual-path LSTM architecture for sequential modeling. The network includes two modeling paths, (a) intra-segment (local: path 1) and (b) inter-segment (global: path 2).

and numerical stability value respectively. In (1), $E\{\cdot\}$, $\text{Var}\{\cdot\}$, and $\odot$ denote expectation value, variance, and the Hadamard product operator respectively. The input feature map is transferred to the output layer through a skip connection, allowing for information flow from the initial layer to the last layer. As such, the output of the local modeling path, shown by O_l , is a summation of the normalized matrix and input feature map (N_l and S_l respectively), which serves as the input to the global modeling path.

Fig. 3(b) illustrates the global modeling path in the DP-LSTM network. The structure of this path is similar to that of the local modeling path except for the direction of the Bi-LSTM module which is applied to each of the L sequences separately. The output of the Bi-LSTM, fully-connected, and normalization layers are shown by B_g , F_g , and N_g respectively. The output of this layer is a summation of the input feature and the normalized matrix (N_g and S_g respectively). As a result of the local and global modeling paths, the information of the input signal is extracted for further processing.

2) Abdominal ECG Denoising: As mentioned earlier, abdominal recordings are contaminated by noise components. Generative adversarial networks (GAN), initially proposed by Ian Goodfellow et al. [65], have been widely used in a variety of applications, an instance of which is ECG signal denoising [66], biomedical image synthesis [67], and speech synthesis [68]. Fig. 4 shows the developed conditional least-square generative adversarial network (LSCGAN) for noise mitigation. This network is composed of two sub-modules, a DP-LSTM generator and a discriminator based on residual blocks, as illustrated in Fig. 4(a) and described below:

Generator: As shown in Fig. 4(a), the generator consists of a 1-D convolutional layer with 16 filters, a DP-LSTM module, and a depth-wise convolution. The 1-D convolutional layer takes a single-channel noisy abdominal ECG signal $x_c \in \mathbb{R}^{1 \times T}$, and transforms it into $x_c \in \mathbb{R}^{M \times T}$, where $M=16$ and T shows the length of the signal. The DP-LSTM block then segments the signal and extracts its temporal features ($O_l \in \mathbb{R}^{M \times K \times L}$), which are then expanded back to their original dimensions ($\mathbb{R}^{M \times T}$). The resulting M channels are transformed to a denoised AECG signal, i.e., $G(x_c)$, using a depth-wise convolution (Conv3, 1).

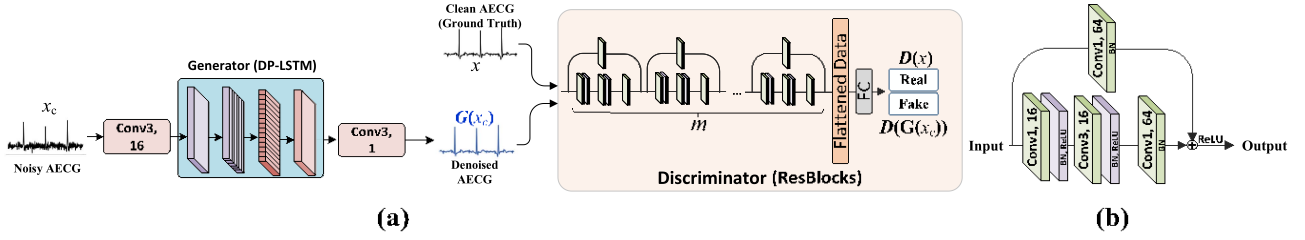


Fig. 4. (a) The CGAN-based denoising network consisting of a dual-path LSTM (DP-LSTM) generator and a discriminator based upon residual blocks. (b) A residual block structure.

Discriminator: The discriminator consists of m 1-D residual blocks (ResBlocks). Each residual block incorporates three convolutional layers accompanied by a skip connection as shown in Fig. 4(b), where the rectified linear unit (ReLU) and batch normalization (BN) are employed to add non-linearity and stabilize the training procedure, respectively. The ResBlocks are followed by a fully-connected (FC) layer to discriminate between real (x) and synthetic data ($G(x_c)$).

Training: In GAN architectures, the generator and discriminator networks are trained simultaneously. Given that x_c and x are sampled from the noisy abdominal ECG distribution ($p_{data}(x_c)$) and the clean abdominal ECG distribution ($p_{data}(x)$), the generator aims to map x_c to x through adversarial training. On the other hand, the discriminator is supposed to classify samples from $p_{data}(x)$ as real and samples from $p_{data}(x_c)$, i.e., $G(x_c)$, as fake. As such, the generator tries to imitate the training data and deceive the discriminator, whereas the discriminator is trained to distinguish fake data. The training procedure is conducted in two steps. In the first step, the discriminator is trained to classify x as real data ($p_{data}(x)$). In the second step, the generator learns to imitate $p_{data}(x)$ such that the discriminator classifies $G(x_c)$ as real data. Hence, two loss terms corresponding to the discriminator and generator networks are defined as (2) and (3), respectively.

$$\min_D L_{LSGAN}(D) = \frac{1}{2} E_{x \sim p_{data}(x)} (D(x) - 1)^2 + \frac{1}{2} E_{x_c \sim p_{data}(x_c)} D(G(x_c))^2, \quad (2)$$

$$\min_G L_{LSGAN}(G) = \frac{1}{2} E_{x_c \sim p_{data}(x_c)} (D(G(x_c)) - 1)^2 + \lambda k G(x_c) - x k_1. \quad (3)$$

In (3), the term $\lambda k G(x_c) - x k_1$ is used to minimize the distance (L1-norm) between $G(x_c)$ and x . This term is fused with the adversarial loss term with a coefficient of λ which was found to be 0.01 based on our empirical results. The whole process of training repeats until the generator learns the distribution of clean abdominal ECG signals. Therefore, the generator would be capable of denoising abdominal ECG signals.

3) Mask-Based Source Separation: Mask-based source separation is illustrated in Fig. 5. This step receives the denoised abdominal ECG signal from the denoising network, extracts features, and eventually separates fetal and maternal components from each other.

Sparse Feature Extraction: ECG signals hold sparse nature composed of regular activities (P, QRS, and T waveforms) as well as periods of inactivity [69]. To assure accurate source separation, we need to extract fine and coarse features of abdominal ECG signals. To this end, we utilize an inception module with dimensionality reduction as codenamed by [70] for the GoogleNet architecture and depicted in Fig. 5(a). As such, the denoised abdominal ECG of length T is segmented into a matrix of dimensions $K \times 1 \times L$. This matrix undergoes the inception module with four paths. These paths leverage convolutional layers with filter sizes of 1, 3, and 5 as well as a pooling layer used to reinforce invariance to signal distortion. The number of paths (the limitation on filter sizes) will be investigated in Section IV. The outputs of all paths are then concatenated to build a feature map of dimensions $K \times M \times L$. The logic behind such feature extraction is to transfer the abdominal ECG signal into a higher level of abstraction to simplify the separation procedure. The output of the inception module serves as the input to the dual-path masking block as illustrated in Fig. 5(b).

Dual-path Masking: To separate fetal and maternal features from each other, their respective features are masked in high levels of abstraction. To this end, a series of DP-LSTM modules are cascaded to extract masking functions from the sparse feature map as shown in Fig. 5(b). The output of the DP-LSTM blocks is a matrix with dimension $K \times M \times L$. To extract maternal and fetal components, two masking functions are required to be multiplied by the sparse feature map which is of dimension $K \times M \times L$. These masking functions are built from the output of DP-LSTM series using a 2D convolutional layer with $2M$ filters followed by a sigmoid function. As a result, a mask-ing function with a size of $K \times 2M \times L$ is obtained which is then split into two masking maps of sizes $K \times M \times L$. Next, each masking function is multiplied by the sparse feature map separately to discriminate features representing fetal and maternal components. Each of the acquired FECG and MECG components consist of M features along the second dimension. To transform the features back into the time-domain, the M features should be consolidated into a single feature. As such, two 2-D depth-wise convolutional layers using a single feature map are utilized to generate FECG and MECG signal segments. Finally, the segments are assembled to build the original FECG and MECG signals.

Training: The objective of training the source separation network is to maximize the signal-to-noise ratio (SNR). As the temporal variation of the signal, rather than its order of magnitude, holds information about the cardiac cycles, we optimize

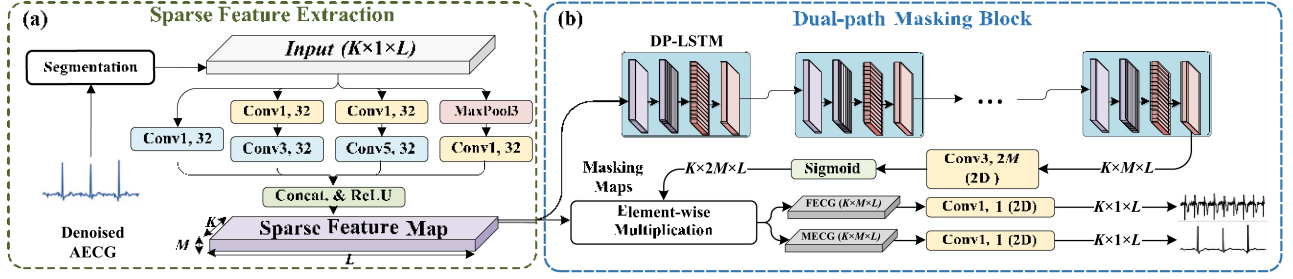


Fig. 5. Source separation network including (a) sparse feature extraction and (b) dual-path masking blocks. The input to the network is a denoised abdominal ECG which undergoes feature extraction and masking. The fetal and maternal ECG signals constitute the output of the network.

the scale-invariant SNR (SI-SNR) function which is robust to the magnitude scales of the estimated signal [71]. For a ground-truth signal x and its estimated counterpart $\hat{x}$, SI-SNR is defined as follows:

$$\text{SI-SNR} = 10 \log \frac{kx_{\text{target}}k^2}{ke_{\text{noise}}k^2}, \quad (4)$$

where

$$x_{\text{target}} = \frac{h\hat{x}, x}{k\hat{x}k^2}, \quad e_{\text{noise}} = \hat{x} - x_{\text{target}}, \quad (5)$$

and k and h denote the Euclidean norm and scalar product respectively. As such, the loss function for training the source separation network is defined as the negative value of the SI-SNR, as mentioned below:

$$\min L_{\text{SI-SNR}} = -10 \log \frac{kx_{\text{target}}k^2}{ke_{\text{noise}}k^2}. \quad (6)$$

During training, the error between fetal ECG (s_{FECG}) and maternal ECG (s_{MECG}) signals and their estimated counterparts [$\hat{s}_{\text{FECG}}$, $\hat{s}_{\text{MECG}}$] are achieved using the $L_{\text{SI-SNR}}$ loss function. However, since the model estimates the masking maps simultaneously using the same input abdominal ECG, the order of the output signals is unknown in advance. As such, the output can have either [$\hat{s}_{\text{FECG}}$, $\hat{s}_{\text{MECG}}$] or [$\hat{s}_{\text{MECG}}$, $\hat{s}_{\text{FECG}}$] arrangements. This issue causes training failure. To address this problem, we employ the permutation invariant training (PIT) strategy [72], where the two possible permutations for FECG and MECG are considered. As such, at every training iteration, the average errors for both orders are evaluated, and the order with the least average loss is selected to back-propagate the error and update the weights of the network.

D. Data Augmentation

Neural networks tend to suffer from overfitting due to the limited amount of training data [73]. This phenomenon leads the network to excessively fit the training data, restraining the generalizability of the network to unseen data [74]. On the other hand, neural networks can benefit from data augmentation which aims at generating synthetic data by transforming existing samples into their modified versions [75]. The augmented data space enforces stability in training by reducing the variance [75]. In this study, we employ three techniques, namely jittering,

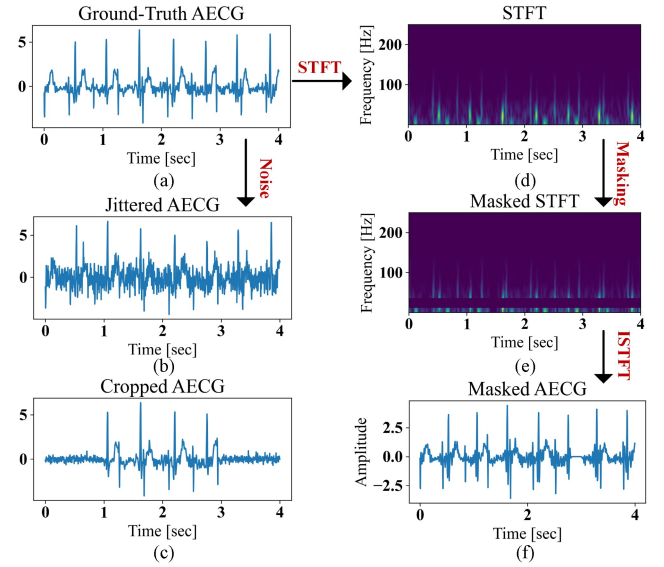


Fig. 6. Data augmentation techniques applied on (a) abdominal ECG (AECG) from FECGSYNDB, (b) jittered AECG, (c) cropped AECG, (d) STFT of AECG, (e) masked STFT of AECG, and (f) masked ECG.

cropping, and time-frequency masking, to augment the training dataset. Fig. 6 shows the ground-truth signal and the augmented samples.

1) Jittering: An effective data augmentation method is jittering wherein zero-mean white Gaussian noise, i.e., $N(0, \sigma^2)$, is added to training samples. Furthermore, pregnancy-specific noise such as uterine contractions and fetal movement are combined with abdominal signals to simulate realistic scenarios. An example of jittering is illustrated in Fig. 6(b), where an abdominal ECG from FECGSYNDB is contaminated by noise components. This practice helps mitigate the time series drift caused by unseen data [76]. In this study, the standard deviation of the normal distribution is selected such that the noise power falls 3 dB below the energy of FECG QRS complexes.

2) Cropping: Random cropping aims to create a random subset of original abdominal ECG recordings by cropping a random section of the signal. QRS complexes in an abdominal recording are not always wholly visible in the signal segment which might hinder the network to learn such cases. Random

cropping deals with these cases, helping the network to learn the data distribution more comprehensively, and enhance generalizability. Hence, the Beta distribution whose superiority for data augmentation over the uniform distribution has been demonstrated [77], was utilized to specify the cropping width, w . This value is selected as below:

$$w = bw^0Tc, w^0 \sim \text{Beta}(\beta, \beta), \quad (7)$$

where $\beta \in (0, +\infty)$, and $b.c$ and T denote the floor function and the original segment length, respectively. As such, a window of length w centered in the middle of the original segment is cropped and the held-out samples are padded using a zero-mean normal distribution with SNR=10 dB. An example of a 50% random cropping is presented in Fig. 6(c), where a fetal QRS peak is located on the boundary.

3) Time-Frequency Masking: In this technique, the time-frequency representation of each segment is obtained through the short-time Fourier transform (STFT) as shown in Fig. 6(d). Then, a number of randomly selected bins are set to zero, which builds a new time-frequency representation as illustrated in Fig. 6(e). Finally, the inverse STFT (ISTFT) is applied to the generated time-frequency representation to transform it back into the time domain. This technique can attenuate certain spectral features in ECG (such as T-waves as shown in Fig. 6(f)) as well as the temporal evolution of the signal (e.g., around 2.9 s in Fig. 6(f)). As such, time-frequency masking reinforces the network with learning a more general distribution. It should be noted that random bins are chosen from frequency bins below 50 Hz due to the concentration of ECG information in this region [78].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, training procedures and experimental results are discussed. Moreover, an ablation study is presented to investigate the network configuration.

A. Training Procedure

The dual-path source separation (DPSS) represents a supervised learning configuration. It thus demands access to noisy abdominal ECG, clean abdominal ECG, fetal ECG, and maternal ECG signals. The only dataset which provides all the required signals is FECGSYNDB, from which the data of ten different pregnancies were used for training. The training procedure was conducted based on leave-subject-out cross-validation. As such, the data of 9 subjects were used for training and the performance was evaluated on the held-out subject. For the sake of computational simplicity, the denoising and source separation networks were trained separately. To further evaluate the performance of the method, the trained model for FECGSYNDB was tested on ADFECGDB(P), ADFECGDB(L), and NIFECGC datasets as well.

We trained our model for 200 epochs on abdominal ECG segments. The Adam optimization algorithm [79] was used with an initial learning rate of 0.001, which was reduced by a factor of 0.98 for every five epochs without performance improvement.

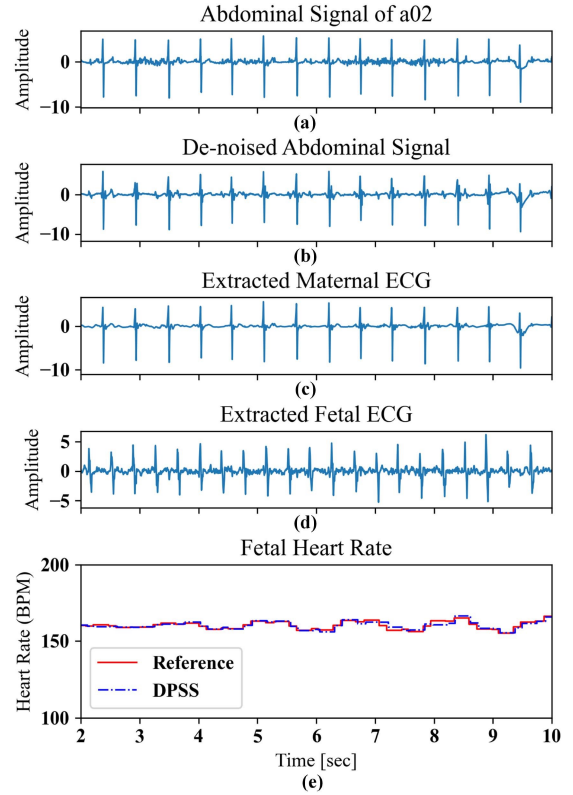


Fig. 7. An 8-second test segment reconstructed by overlapping the segments of a02 from the NIFECGC dataset. (a) Noisy abdominal recording, (b) de-noised abdominal ECG, (c) extracted maternal ECG, (d) extracted fetal ECG, and (e) estimated fetal heart rate.

The model was implemented on an NVIDIA GeForce RTX 2070 with a batch size of 32.

Fig. 7(a) illustrates an 8-second noisy abdominal ECG segment of recording a02 from NIFECGC. This signal was segmented into overlapping 4-second chunks, denoised, source-separated, and eventually reconstructed by merging the segments. As such, Fig. 7(b), (c), (d), and (e) depict the merged output of the denoiser network, maternal ECG, fetal ECG, and estimated FHR, respectively. These figures visually demonstrate the effectiveness of DPSS in fetal ECG extraction. As observed in Fig. 7(d), the peaks corresponding to maternal QRS waveforms have been majorly attenuated, allowing for the detection of the FECG R-peaks. In parallel, Fig. 7(e) illustrates high consistency between the estimated FHR values by DPSS with their respective ground-truth values. Further results on the performance of DPSS are provided in supplementary materials. In Section IV-C, the performance is analyzed numerically.

B. Evaluation Metrics

1) Statistical Analyses of Fetal QRS Detection: Statistical analyses were conducted to evaluate the performance of the DPSS network and compare it with the state-of-the-art methods of single-channel fetal ECG extraction. In order to compare the

extracted fetal ECG segments with their ground-truth counterparts, the modified version of the Pan-Tompkins method for fetal QRS detection was employed [51]. Following the detection of R-peaks in both reference and extracted FECG signals, sensitivity (Sen), positive predictive value (PPV), accuracy (Acc), and F1-score (F1) were obtained as follows:

$$\text{Sen} = 100 \times \frac{TP}{TP + FN}, \quad (8)$$

$$\text{PPV} = 100 \times \frac{TP}{TP + FP}, \quad (9)$$

$$\text{Acc} = 100 \times \frac{TP}{TP + FP + FN}, \quad (10)$$

$$\text{F1} = \frac{2 \times PP \times \text{Sen}}{PP + \text{Sen}}, \quad (11)$$

where TP , FP , and FN represent true positives (correctly detected FECG QRS complexes), false positives (falsely detected non-existent FECG QRS complexes), and false negatives (falsely missed existent FECG QRS complexes), respectively. According to [14], a detected QRS is considered a true positive if it is within 50 ms from the reference R-peak. Additionally, extracted FECG segments are assessed by the FHR precision, which is defined as the percentage of segments wherein the extracted FHR is within ± 5 beats-per-minute (BPM) of the reference FHR.

2) Clustering-Based Source Separation Analyses: As mentioned in Section III, DPSS generates two $K \times M \times L$ masking maps to separate fetal and maternal ECG signals. These masking maps are obtained by learning the distributions of MEG and FECG signals from the input AECG signals. A well-trained network is meant to acquire two separable distributions in the embedding space, corresponding to fetal and maternal ECG components. In this work, we quantify the quality of separation by the Davies-Bouldin index defined below:

$$\text{DB} = \frac{\sigma_M + \sigma_F}{2d(c_M, c_F)}, \quad (12)$$

where c_M and c_F represent the centroids of the maternal and fetal clusters, respectively, and σ_M and σ_F show the average distance of all maps in maternal and fetal clusters to their respective centroids respectively. Also, $d(c_M, c_F)$ represent the distance between the centroids of maternal and fetal clusters. This metric implies the ratio of intra-cluster distances to those of inter-clusters. As such, lower values of DB indicate better clustering. In this work, we conduct a two-class K-means algorithm to cluster masking maps. We then find the Davies-Bouldin index using the ground-truth labels (fetal or maternal).

Furthermore, we compute the Jaccard index to assess the similarity between the clustered group and its corresponding labels (either maternal or fetal). The Jaccard index is defined as follows:

$$\text{Jaccard}(A, A_i) = \frac{|A \cap A_i|}{|A \cup A_i|} = \frac{tp}{tp + fp + fn}, \quad (13)$$

where tp , fp , and fn signify the numbers of truly-included elements, falsely-included elements, and falsely-excluded elements in a cluster, respectively. To compute the Jaccard metric, each

cluster is assigned to either maternal or fetal based on the labels of the majority of the samples. The Jaccard metric is the average of the Jaccard values of the maternal and fetal clusters.

The last parameter is the purity index. This parameter describes the extent to which clusters contain their dominant elements. For each cluster, the dominant elements are counted and summed over all clusters. The sum value is then divided by the total number of elements.

C. Model Performance

Table I summarizes the statistical analysis of the performance of the dual-path source separation (DPSS) framework on the FECGSYNDB dataset. The results in Table I correspond to a DPSS architecture with three dual-path LSTMs and six feature extraction filters with lengths 1, 3, 5, 7, 9, and 11. The length of the sub-segments was set to $L = 100$ ms for temporal modeling in DPSS. Per each held-out subject, the model is trained on the augmented data of 9 other subjects, which includes 3,432,960 4-second segments of AECG signals. The performance is then reported by the sensitivity (Sen), positive predictive value (PPV), accuracy (Acc), F1-score (F1), and the maximum SI-SNR achieved during the training for the 190,720 test segments. As mentioned in the table, the best performance is achieved for subject 8 with sensitivity, positive predictive value, accuracy, F1-score, and SI-SNR of 99.09%, 99.61%, 98.71%, 99.35%, and +15.94 dB respectively. The weakest source separation performance belongs to subject 2 with 98.24% and +13.35 dB F1-score and SI-SNR respectively. However, the average ($\pm$ standard deviation) F1-score and SI-SNR suggest 99.03% (± 0.39) agreement with the reference QRS complexes and +14.92 (± 1.02) dB signal-to-noise ratio in the output, respectively. Such a decent performance is expected as the test set is sampled from the same dataset and thus the same data distribution as the training set.

To further evaluate the generalizability of DPSS, the trained model is tested on the ADFECGDB(P), ADFECGDB(G), and NIFECGC datasets with 2,999, 749, and 149 segments per channel respectively to reconstruct FECG signals, and their statistical results are summarized in Table II. According to Table II, the labor dataset achieves the highest F1-score (98.08%) and FHR precision (89.22%), while NIFECGC obtains the weakest performance with 83.86% of FHR precision and 95.3% of F1-score. ADFECGDB, which denotes the average result of labor and pregnancy dataset, suggests 97.7% and 88.61% F1-score and FHR precision, respectively. The slight difference of statistical results of ADFECGDB and NIFECGC can be explained by stronger fetal QRS peaks in ADFECGDB dataset. The last row of Table II signifies the average performance of DPSS on real datasets, i.e., NIFECGC and ADFECGDB, with sensitivity of 86.23%, positive predictive value of 95.75%, accuracy of 97.29%, and F1-score of 93.43%, demonstrating the robustness of the method.

D. Clustering-Based Evaluation

As introduced in Section IV-C, the purity, Jaccard, and Davies-Bouldin indices are used to quantify and evaluate the

TABLE I

CROSS-VALIDATION PERFORMANCE FOR FECGSYNDB DATASET

Subject	Sen (%)	PPV (%)	Acc (%)	F1 (%)	SI-SNR (dB)
Subject 1	98.78	99.77	98.55	99.27	14.62
Subject 2	97.43	99.07	96.55	98.24	13.35
Subject 3	98.95	99.07	98.04	99.01	16.91
Subject 4	98.82	99.54	98.37	99.18	15.36
Subject 5	97.54	99.21	96.79	98.37	14.32
Subject 6	99.02	99.6	98.62	99.31	15.02
Subject 7	98.61	99.71	98.32	99.16	14.67
Subject 8	99.09	99.61	98.71	99.35	15.94
Subject 9	98.65	99.83	98.48	99.24	13.88
Subject 10	98.58	99.76	98.35	99.17	15.13
Average	98.55 (± 0.59)	99.52 (± 0.29)	98.08 (± 0.77)	99.03 (± 0.39)	14.92 (± 1.02)

TABLE II

PERFORMANCE EVALUATION OF DPSS ON THE ADFECGDB AND NIFECGC DATASETS

Dataset	FHR Precision (%)	Sen (%)	PPV (%)	Acc (%)	F1 (%)
ADFECDGB(P)	87.88	96.5	97.98	94.67	97.24
ADFECDGB(L)	89.22	98.0	98.18	96.25	98.08
ADFECDGB	88.61	97.3	98.09	95.53	97.7
NIFECGC	83.86	94.2	96.5	91.34	95.3
Average of two datasets	86.23	95.75	97.29	93.43	96.5

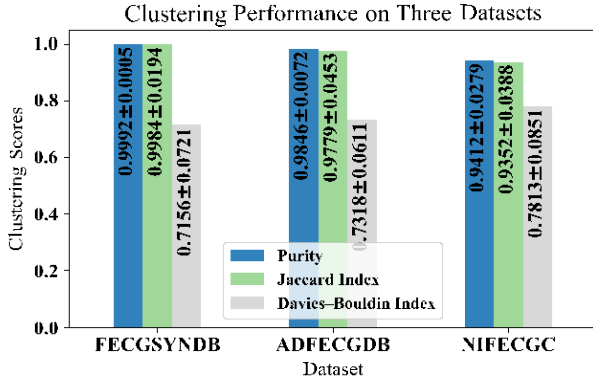


Fig. 8. Clustering performance for FECGSYNDB, ADFECGDB, and NIFECGC datasets based on the purity, Jaccard, and Davies-Bouldin indices.

quality of source separation in DPSS. As shown in Fig. 5, two masking maps (each of size $K \times M \times L$) are produced to separate the fetal and maternal components of a 4-second mixture AECG signal. For ideal source separation, we expect the maternal and fetal masking maps of all segments to form two separate clusters in the $R^{K \times M \times L}$ space. To investigate this phenomenon, the aforementioned metrics can be leveraged provided that each $K \times M \times L$ masking map represents a sample in the $R^{K \times M \times L}$ space. However, the computation of these metrics is a computationally intensive task, motivating the need for dimensionality reduction in advance. For this purpose, we use the t-distributed stochastic neighbor embedding (t-SNE) [81]. As such, clustering-based analyses were conducted on the reduced-dimensionality samples, and the results are summarized in Fig. 8. As described in this figure, the highest and lowest clustering performances were achieved for the FECGSYNDB and NIFECGC datasets with average ($\pm$ standard deviation) purity indices of 0.9992 ($\pm$ 0.0005) and 0.9412 ($\pm$ 0.0279) respectively, whereas

TABLE III

ABLATION STUDY BASED ON THE PERFORMANCE FOR FETAL ECG EXTRACTION FOR FECGSYNDB

Architecture	F1-score (%)
1. Baseline (1 DP)	61.76 (± 1.52)
2. Baseline (1 DP) + DN	84.21 (± 1.06)
3. Baseline (2 DP) + DN	84.10 (± 1.17)
4. Baseline (3 DP) + DN	85.13 (± 0.96)
5. Baseline (4 DP) + DN	82.1 (± 0.82)
6. Baseline (3 DP + SF3) + DN	88.73 (± 0.45)
7. Baseline (3 DP + SF4) + DN	90.16 (± 0.48)
8. Baseline (3 DP + SF5) + DN	93.07 (± 0.68)
9. Baseline (3 DP + SF6) + DN	94.68 (± 0.39)
10. Baseline (3 DP + SF7) + DN	92.43 (± 0.93)
11. Baseline (3 DP + SF8) + DN	91.77 (± 0.81)
12. Baseline (3 DP + SF6) + DN + DA	99.03 (± 0.39)

DP: Dual-path LSTM, DN: denoising, SF: sparse feature extraction, and DA: data augmentation.

a purity index of 0.9846 ($\pm$ 0.0072) was achieved for ADFECGDB, implying high concentration of the dominant samples (either maternal or fetal) in their respective cluster. Similarly, the Jaccard values of 0.9984 ($\pm$ 0.0194), 0.9779 ($\pm$ 0.0453), and 0.9352 ($\pm$ 0.0388) respectively suggest high similarities between FECGSYNDB, ADFECGDB, and NIFECGC datasets and their respective ground-truth labels. Moreover, the Davies-Bouldin (DB) values signify the separability of the clusters for FECGSYNDB, ADFECGDB, and NIFECGC by 0.7156 ($\pm$ 0.0721), 0.7318 ($\pm$ 0.0611), and 0.7813 ($\pm$ 0.0851) respectively. It is to be noted that the lower the Davies-Bouldin metric, the higher the separability. According to the achieved Davies-Bouldin values, the intra-cluster distances are smaller than their respective inter-cluster distances, suggesting separable fetal and maternal masking maps.

E. Ablation Study

A common practice to better generalize a neural network to unseen data (real datasets) is by an ablation study on the constituent parts of the training procedure. In this study, we investigate the impacts of the denoising module (DN), the number of dual-path LSTM modules (DP), the sparse feature extraction mechanism (SF), the data augmentation (DA), and the lengths of segments and sub-segments on the performance of the DPSS network. The ablation study is conducted for each fold of FECGSYNDB and averaged over all 10 subjects to maximize the network F1-score.

Comparing architectures 1 and 2 in Table III illustrates the impact of the denoising network on the F1-score (61.76% vs. 84.21%), suggesting a 22.45% improvement as a result of incorporating the denoising network. According to architectures 1-5, including 3 dual-path LSTM (DP) modules leads to the superior performance among all similar architectures with 1, 2, 3, and 4 DP module(s). An F1-score of 85.13% demonstrates the superiority over the runner-up architecture with 1 DP module and F1-score of 84.21%. The weaker performance in architecture 5 compared to Baseline (1 DP) + DN can be related to the higher complexity of the model, which causes the network to overfit the data.

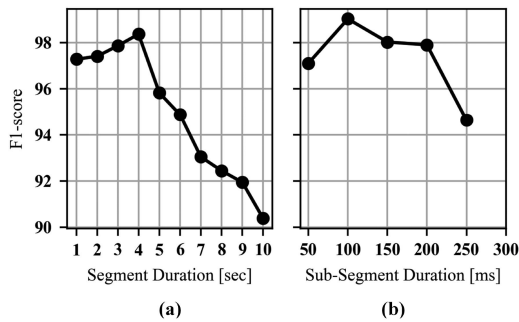


Fig. 9. Impacts of (a) segment duration and (b) sub-segment duration on the F1-score obtained for the detection of fetal QRS complexes in FECGSYNDB.

A comparison conducted among architectures 6-11 implies that the model with 6-filter sparse feature extraction, which is referred to as SF6, offers the highest F1-score (94.68%). It should be noted that architecture Baseline (3 DP + SF6) + DN leverages filters with sizes of 1, 3, 5, 7, 9, and 13. Intuitively, employing filters with different sizes allows for extracting both fine and coarse features which help the neural network learn the distribution mode deeply. Nevertheless, increasing the number of filters results in more complex architectures that might overfit the distribution of the data. Two examples of overfitting are observed in Table III for architecture 10 with SF7 and architecture 11 with SF8 with F1-scores of 92.43% and 91.77%, respectively.

A comparison between architectures 9 and 12 signifies the importance of data augmentation (DA). For this part, the dataset is augmented to achieve a twice larger dataset than the original size (3,432,960 vs. 1,716,480 segments). As reported in Table III, the Baseline (3 DP + SF6) + DN + DA outperforms the model without data augmentation (Baseline (3 DP + SF6) + DN) by 4.35% F1-score.

Finally, the impacts of the segment length (T) and sub-segment length (L) are investigated in Figs. 9(a) and (b) respectively. According to Fig. 9(a), the model is trained for segment lengths within the range of 1–10 seconds, where the highest average F1-score of 98.37 % is achieved for a segment length of 4 seconds. As can be seen in this figure, the average F1-score decreases from 98.37% to 90.39% when moving from 4 seconds to 10 seconds. Fig. 9(b) depicts the average F1-score in terms of sub-segment duration, where 100 ms suggests a higher F1-score compared to other durations (99.03% vs. 97.11%, 98.03%, 97.91%, and 94.65% for 50 ms, 150 ms, 200 ms, and 250 ms respectively).

F. Comparative Study on Fetal QRS Detection and FHR Estimation

In Table IV, DPSS is compared with state-of-the-art single-channel fetal ECG extraction methods. As summarized in Table IV, an F1-score of 95.3% on the NIFECGC dataset suggests the superiority of DPSS over the previous works except [9], [18], [46], and [47]. Authors in [9] and [18] have used all 4 abdominal channels, whereas DPSS requires only a single channel to reconstruct FECG signal. Performance achieved in [46] and [47] represent 17 and 27 subjects, respectively, whereas our work is evaluated on all 69 subjects with correct

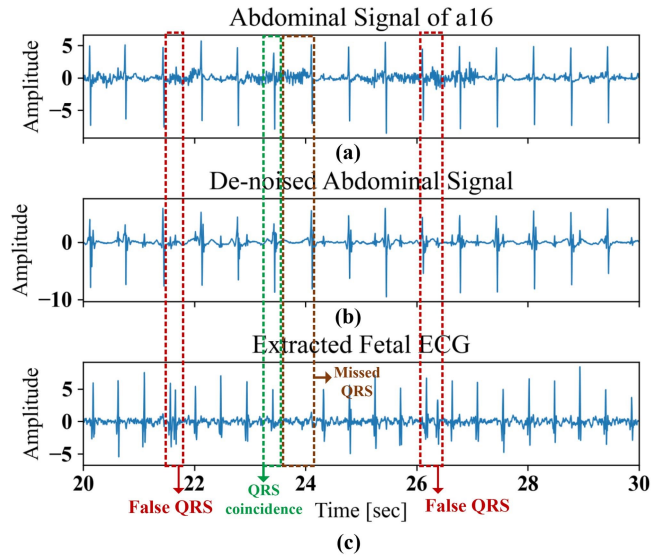


Fig. 10. Examples of false detection of fetal QRS, successful detection of coincided fetal and maternal QRS peaks, and missed fetal QRS. (a) abdominal signal of a16 from NIFECG, (b) de-noised abdominal signal, and (c) extracted fetal ECG signal.

signal annotations. In terms of FHR precision, DPSS suggests 88.61% and 83.86% for ADFECGDB and NIFECGC, respectively, outperforming all previous works which have reported on FHR precision values. In addition, DPSS shows promising results on the ADFECGDB dataset with an F1-score of 97.7% on 22 signals, which is a superior estimation to those conventional methods except [49], [47], [46], and [33] with 99.7%, 98.04%, 99.27%, and 99.44% respectively. Although DPSS provides slightly weaker performance, it should be noted that we have used the extended version of ADFECGDB (labor and pregnancy) in contrast to the literature. In general, DPSS suggests quite an acceptable performance in comparison to literature despite its low computational demands, which will be indicated in the following section.

V. DISCUSSION

As demonstrated in Section IV, DPSS is capable of fetal ECG extraction from abdominal recordings. A distinctive characteristic of DPSS is its robustness against coincided fetal and maternal QRS peaks, which often inhibits conventional methods from detecting fetal peaks. This is explained by a GAN-based denoiser which cancels noise components by learning the distribution of the abdominal signals. Furthermore, the results of this study demonstrate that masking functions, directly extracted from the abdominal signal, allow for accurately separating fetal components in time domain. Fig. 10 provides a 10-second example of recording a016 from the NIFECGC dataset. As can be observed in Fig. 10(c), the green rectangle signifies an example of fetal and maternal QRS peaks, where masking functions have successfully separated the sources. Additionally, a significant finding of this study is fetal QRS amplification by denoiser. A comparison between fetal QRS peaks in Figs 10(a) and (b) indicates that fetal QRS peaks are amplified in terms of amplitude as a result of the denoising stage.

TABLE IV
COMPARATIVE EVALUATION FOR THE DETECTION OF FETAL QRS COMPLEXES AND FHR ESTIMATION VS. STATE-OF-THE-ART METHODS

Method	Dataset	Channels	Recordings	Year	FHR Precision (%)	Sen (%)	PPV (%)	Acc (%)	F1 (%)
TS (Cerutti et al.) [30]	NIFECG	1	14	1986	57.9	91.2	90.5	NA	90.8
	Private				68.1	86.8	85.3		86
TS + PCA (Kanjilal et al.) [28]	NIFECG	1	14	1997	68.7	94.7	96	NA	95.4
	Private				73.9	89.9	88.8		89.3
LMS (Behar et al.) [24]	NIFECG	1	14	2014	69.3	95.8	95	NA	95.4
	Private				73.7	89.3	86.5		87.9
RLS (Behar et al.) [24]	NIFECG	1	14	2014	70.6	96.2	95.6	NA	95.9
	Private				74.5	89.7	86.8		88.2
ESN (Behar et al.) [24]	NIFECG	1	14	2014	72.2	96.8	97.2	NA	97
	Private				78.7	91.4	88.9		90.2
Behar et al. [9]	NIFECG	4	69	2014	NA	95.9	96	NA	96
Varanini et al. [18]	NIFECG	4	69	2014	NA	99.1	98.9	NA	98.99
SVD + SW (Zhang et al.) [33]	ADFECEGDB	1	2	2017	NA	99.36	99.52	98.89	99.44
	Private	1	1			98.31	98.86	97.21	98.58
STFT + NMF (Gurve et al.) [45]	ADFECEGDB	1	5	2019	NA	95.3	94.6	NA	94.8
	NIFECG		60			NA	NA		83.8
Mohebbian et al. [80]	NIFECG	1	14	2020	NA	95.3	97.1	NA	96.1
TF + Shannon (Krupa et al.) [46]	ADFECEGDB	1	13	2022	NA	99.37	99.38	98.55	99.27
	NIFECG		26			98.61	98.72	97.37	98.67
Clustering (Zhang et al.) [47]	ADFECEGDB	1	11	2018	NA	97.5	98.58	NA	98.04
	NIFECG		64			97.51	98.63		98.07
RCED-NET (Zhong et al.) [48]	ADFECEGDB	1	9	2019	NA	96.06	92.25	NA	94.1
	NIFECG		40			92.6	94.68		93.62
CycleGAN (Mohebbian et al.) [49]	ADFECEGDB	1	5	2021	NA	99.4	99.6	NA	99.7
	NIFECG		14			96.8	97.2		97.9
STFT + GAN (Zhong et al.) [50]	NIFECG	1	69	2021	NA	93.9	93.9	NA	94.7
	ADFECEGDB		5			90.32	89.79		90.05
DPSS (Proposed)	NIFECG	1	80	2022	NA	92.37	93.69	NA	93.02
	ADFECEGDB(P)		10			87.88	96.5	97.98	97.24
	ADFECEGDB(L)		12			89.22	98	98.18	96.25
	ADFECEGDB		22			88.61	97.3	98.09	95.53
	NIFECG		69			83.86	94.2	96.5	91.34

TS: template subtraction, PCA: principal component analysis, LMS: least mean-square, RLS: recursive mean-square, ESN: echo state network, SVD: singular value decomposition, SW: smooth window, STFT: short-time Fourier transform, NMF: non-negative matrix factorization, TF: time-frequency, RCED-NET: residual convolutional encoder-decoder network, GAN: generative adversarial network, and DPSS: dual-path source separation.

Despite the promising results achieved on real datasets, the performance of DPSS is limited when the GAN-based denoiser enhances artifacts that resemble fetal QRS peaks. As demonstrated in the red panels of Fig. 10, DPSS might detect noise components as fetal QRS peaks, leading to false positive cases. Fetal QRS peaks might also be missed by the denoiser as a result of high-energy noise, an example of which is illustrated in the brown panel of Fig. 10(a). According to its counterpart in Fig. 10(b), the whole segment is smoothed, eliminating fetal QRS waveform and increasing false negatives. Another finding about the extracted fetal ECG signals is the oscillatory behavior in between consecutive QRS peaks which contaminates fetal QRS P and T peaks. According to our observation, this phenomenon further degrades the quality of fetal ECG signals when it is accompanied by the remnant of maternal QRS peaks.

To have a real-time algorithm, the processing time of a signal segment should be less than the segment length. In order to investigate the real-time practicality of DPSS, we transferred the trained DPSS network to an Intel processing unit of a computer with a 3.6 GHz clock speed. It was observed that the processing unit for a 4-second abdominal signal was 0.52 s, as illustrated in Fig. 11. As demonstrated in this figure, the processing time was also calculated for off-the-shelf processors, such as Broadcom BCM2711 (in Raspberry Pi), Qualcomm Snapdragon S1 (in smartphones), and Apple S7 (in Apple Watch), the processing times of which are 1.2 s, 1.8 s, and 1 s respectively. These intervals, which are less than 4 seconds, indicate that DPSS

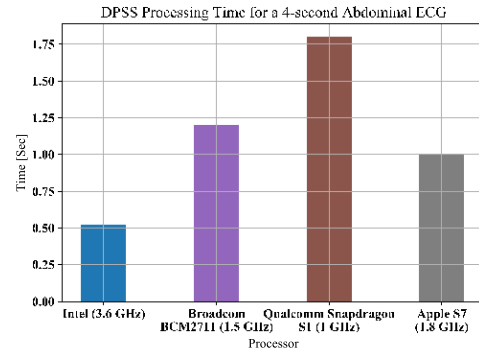


Fig. 11. Processing time of a 4-second abdominal ECG segment on different processors.

holds low computational complexity which makes it a suitable candidate for real-time fetal monitoring systems.

VI. CONCLUSION AND FUTURE WORKS

This study introduces DPSS, a novel method for fetal ECG (FECG) extraction from single-channel abdominal ECG (AECG) recordings. Benefiting from the dual-path long short-term memory (DP-LSTM) mechanism, the DPSS framework estimates FECG and MECG signals by masking raw AECG recordings.

We evaluate the performance of DPSS on the detection of fetal QRS complexes, and demonstrate the F1-scores of 99.03%,

95.3%, and 97.7% on FECGSYNDB with 10 simulated pregnancies, NIFECGC with 69 actual pregnancies, and ADFECGB with 22 actual labor and pregnancies, respectively. The separability of maternal and fetal masking maps is shown by average purity index of 0.975, Jaccard index of 0.9705, and Davies-Bouldin index of 0.7429 respectively. According to the ablation study on the network configuration, the architecture with the denoising module, three dual-path LSTM's, and 6 feature extraction filters, along with data augmentation is found to offer the best performance (F1-score of 99.03%). Additionally, this study demonstrates that the optimum time durations of signal segments and sub-segments for DPSS are 4 seconds and 100 ms respectively.

The proposed source separation framework indicates excellent performance on fetal ECG monitoring for sedentary pregnant women. Nevertheless, part of real-life scenarios involves the pregnant subject moving around such as in walking. These movements affect the signal morphology by introducing additional noise components [82], causing the fetal ECG components to deteriorate. Our future work includes establishing a domain adaptation framework by leveraging generative models such as DicoGAN [83] to mitigate the effects of ambulatory noise. Furthermore, we will consider the separation of maternal and fetal ECG signals when they share the same peak-to-peak intervals over a long period. This phenomenon primarily occurs due to maternal heart rate acceleration or fetal heart rate deceleration, which could potentially degrade the separation capability of the model.

REFERENCES

- [1] R. L. Goldenberg, R. Kirby, and J. Culhane, "Stillbirth: A review," *J. Maternal-Fetal Neonatal Med.*, vol. 16, no. 2, pp. 79–94, 2004.
- [2] UNICEF, "A neglected tragedy, the global burden of stillbirths," 2020. [Online]. Available: <https://data.unicef.org/resources/a-neglected-tragedy-stillbirth-estimates-report/>
- [3] M. Leiva et al., "Fetal cardiac development and hemodynamics in the first trimester," *Ultrasound Obstet. Gynecol.: Official J. Int. Soc. Ultrasound Obstet. Gynecol.*, vol. 14, no. 3, pp. 169–174, 1999.
- [4] R. Martis et al., "Intermittent auscultation (IA) of fetal heart rate in labour for fetal well-being," *Cochrane Database Systematic Rev.*, vol. 2, no. 2, 2017, Art. no. CD008680.
- [5] J. Reinhard et al., "Intrapartum signal quality with external fetal heart rate monitoring: A two way trial of external doppler CTG ultrasound and the abdominal fetal electrocardiogram," *Arch. Gynecol. Obstet.*, vol. 286, no. 5, pp. 1103–1107, 2012.
- [6] I. Voicu et al., "Robust estimation of fetal heart rate from US Doppler signals," *Phys. Procedia*, vol. 3, no. 1, pp. 691–699, 2010.
- [7] FDA, "Avoid fetal 'keepsake' images, heartbeat monitors," 2014. [Online]. Available: <https://www.fda.gov/consumers/consumer-updates/avoid-fetal-keepsake-images-heartbeat-monitors>
- [8] S. Majumder, T. Mondal, and M. J. Deen, "Wearable sensors for remote health monitoring," *Sensors*, vol. 17, no. 1, 2017, Art. no. 130.
- [9] J. Behar, J. Oster, and G. D. Clifford, "Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data," *Physiol. Meas.*, vol. 35, no. 8, 2014, Art. no. 1569.
- [10] F. Kovács et al., "Fetal phonocardiography—Past and future possibilities," *Comput. Methods Prog. Biomed.*, vol. 104, no. 1, pp. 19–25, 2011.
- [11] C. Yang et al., "A pilot study on fetal heart rate extraction from wear-able abdominal inertial sensors," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10773–10781, Nov. 2019.
- [12] A. Shokouhmand et al., "Multi-modal framework for fetal heart rate estimation: Fusion of low-SNR ECG and inertial sensors," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2021, pp. 7166–7169.
- [13] R. Kahankova et al., "A review of signal processing techniques for non-invasive fetal electrocardiography," *IEEE Rev. Biomed. Eng.*, vol. 13, pp. 51–73, 2020.
- [14] J. Behar et al., "A practical guide to non-invasive foetal electrocardiogram extraction and analysis," *Physiol. Meas.*, vol. 37, no. 5, 2016, Art. no. R1.
- [15] R. Martinek et al., "Non-invasive fetal monitoring: A maternal surface ECG electrode placement-based novel approach for optimization of adaptive filter control parameters using the LMS and RLS algorithms," *Sensors*, vol. 17, no. 5, 2017, Art. no. 1154.
- [16] J. A. Lipponen and M. P. Tarvainen, "Principal component model for maternal ECG extraction in fetal QRS detection," *Physiol. Meas.*, vol. 35, no. 8, 2014, Art. no. 1637.
- [17] R. Sameni, C. Jutten, and M. B. Shamsollahi, "Multichannel electrocardiogram decomposition using periodic component analysis," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1935–1940, Aug. 2008.
- [18] M. Varanini et al., "An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG," *Physiol. Meas.*, vol. 35, no. 8, 2014, Art. no. 1607.
- [19] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. San Francisco, CA, USA: Academic Press, 2010.
- [20] E. R. Ferrara and B. Widrow, "Fetal electrocardiogram enhancement by time-sequenced adaptive filtering," *IEEE Trans. Biomed. Eng.*, vol. BME-29, no. 6, pp. 458–460, Jun. 1982.
- [21] M. Shadaydeh, Y. Xiao, and R. K. Ward, "Extraction of fetal ECG using adaptive volterra filters," in *Proc. 16th Eur. Signal Process. Conf.*, 2008, pp. 1–5.
- [22] G. M. Ungureanu et al., "The event synchronous canceller algorithm removes maternal ECG from abdominal signals without affecting the fetal ECG," *Comput. Biol. Med.*, vol. 39, no. 6, pp. 562–567, 2009.
- [23] J. D. K. Abel et al., "Multiple sub-filter adaptive noise canceller for fetal ECG extraction," *Procedia Comput. Sci.*, vol. 165, pp. 182–188, 2019.
- [24] J. Behar et al., "A comparison of single channel fetal ECG extraction methods," *Ann. Biomed. Eng.*, vol. 42, no. 6, pp. 1340–1353, 2014.
- [25] G. D. Clifford et al., "Non-invasive fetal ECG analysis," *Physiol. Meas.*, vol. 35, no. 8, 2014, Art. no. 1521.
- [26] R. Sameni, "Extraction of fetal cardiac signals from an array of maternal abdominal recordings," Ph.D. dissertation, Institut National Polytechnique de Grenoble-INPG, Sharif University, Tehran, Iran, 2008.
- [27] R. Sameni et al., "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2172–2185, Dec. 2007.
- [28] P. P. Kanjilal, S. Palit, and G. Saha, "Fetal ECG extraction from single-channel maternal ECG using singular value decomposition," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 1, pp. 51–59, Jan. 1997.
- [29] S. M. Martens et al., "A robust fetal ECG detection method for abdominal recordings," *Physiol. Meas.*, vol. 28, no. 4, 2007, Art. no. 373.
- [30] S. Cerutti et al., "Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings," *J. Perinatal Med.*, vol. 4, pp. 445–452, 1986.
- [31] M. Ungureanu et al., "Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique," *Biomedizinische Technik*, vol. 52, pp. 56–60, 2007.
- [32] C. Di Maria et al., "Extracting fetal heart beats from maternal abdominal recordings: Selection of the optimal principal components," *Physiol. Meas.*, vol. 35, no. 8, 2014, Art. no. 1649.
- [33] N. Zhang et al., "A novel technique for fetal ECG extraction using single-channel abdominal recording," *Sensors*, vol. 17, no. 3, 2017, Art. no. 457.
- [34] M. Suganthi, S. I. Joy, and B. Sarala, "LabVIEW based non-invasive single channel fetal electrocardiogram extraction," in *Proc. IEEE 1st Int. Conf. Energy, Syst. Inf. Process.*, 2019, pp. 1–5.
- [35] M. Niknazar, B. Rivet, and C. Jutten, "Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1345–1352, May 2013.
- [36] D. Panigrahy, M. Rakshit, and P. Sahu, "An efficient method for fetal ECG extraction from single channel abdominal ECG," in *Proc. Int. Conf. Ind. Instrum. Control*, 2015, pp. 1083–1088.
- [37] D. Panigrahy and P. Sahu, "Extraction of fetal electrocardiogram (ECG) by extended state Kalman filtering and adaptive neuro-fuzzy inference system (ANFIS) based on single channel abdominal recording," *Sadhana*, vol. 40, no. 4, pp. 1091–1104, 2015.
- [38] D. Panigrahy and P. Sahu, "Extraction of fetal ECG signal by an improved method using extended Kalman smoother framework from single channel abdominal ECG signal," *Australas. Phys. Eng. Sci. Med.*, vol. 40, no. 1, pp. 191–207, 2017.

- [39] G. Li et al., "Robust adaptive fetal heart rate estimation for single-channel abdominal ECG recording," in *Proc. 5th Int. Conf. BioMedical Eng. Informat.*, 2012, pp. 587–591.
- [40] P. Gao, E.-C. Chang, and L. Wyse, "Blind separation of fetal ECG from single mixture using SVD and ICA," in *Proc. 4th Int. Conf. Inf. Commun. Signal Process., 4th Pacific Rim Conf. Multimedia. Proc. Joint*, 2003, vol. 3, pp. 1418–1422.
- [41] P. He and X. Chen, "A method for extracting fetal ECG based on EMD-NMF single channel blind source separation algorithm," *Technol. Health Care*, vol. 24, no. s1, pp. S17–S26, 2016.
- [42] P. J. He et al., "Extraction for fetal ECG using single channel blind source separation algorithm based on multi-algorithm fusion," in *Proc. MATEC Web Conf.*, EDP Sciences, 2016, vol. 44, Art. no. 0 1026.
- [43] R. Mujumdar et al., "Principal component analysis (PCA) based single-channel, non-invasive fetal ECG extraction," 2019.
- [44] M. Zhang and G. Wei, "A complete adaptive method for fetal ECG extraction based on single channel," in *Proc. J. Phys.: Conf. Ser.*, 2020, vol. 1621, no. 1, Art. no. 012019.
- [45] D. Gurve and S. Krishnan, "Separation of fetal-ECG from single-channel abdominal eeg using activation scaled non-negative matrix factorization," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 669–680, Mar. 2020.
- [46] A. J. D. Krupa, S. Dhanalakshmi, and R. Kumar, "Joint time-frequency analysis and non-linear estimation for fetal ECG extraction," *Biomed. Signal Process. Control*, vol. 75, 2022, Art. no. 103569.
- [47] E. Castillo et al., "A clustering-based method for single-channel fetal heart rate monitoring," *PLoS One*, vol. 13, no. 6, 2018, Art. no. e0199308.
- [48] W. Zhong et al., "Fetal electrocardiography extraction with residual convolutional encoder-decoder networks," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 4, pp. 1081–1089, 2019.
- [49] M. R. Mohebian et al., "Fetal ECG extraction from maternal ECG using attention-based cycleGAN," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 2, pp. 515–526, Feb. 2022.
- [50] W. Zhong and W. Zhao, "Fetal ECG extraction using short time fourier transform and generative adversarial networks," *Physiol. Meas.*, vol. 42, no. 10, 2021, Art. no. 105011.
- [51] F. Andreotti et al., "An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms," *Physiol. Meas.*, vol. 37, no. 5, 2016, Art. no. 627.
- [52] Y. Luo, Z. Chen, and T. Yoshioka, "Dual-path RNN: Efficient long sequence modeling for time-domain single-channel speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 46–50.
- [53] F. Andreotti, J. Behar, and G. D. Clifford, "Fetal ECG synthetic database," 2016. [Online]. Available: <https://doi.org/10.13026/C21P4T>
- [54] A. Matonia et al., "Fetal electrocardiograms, direct and abdominal with reference heartbeat annotations," *Sci. Data*, vol. 7, no. 1, pp. 1–14, 2020.
- [55] J. Jezewski, et al., "Determination of fetal heart rate from abdominal signals: Evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram," *Biomedizinische Technik/Biomedical Eng.*, vol. 57, no. 5, pp. 383–394, 2012.
- [56] F. Andreotti, J. Behar, and G. D. Clifford, "Abdominal and direct fetal ECG database," 2012. [Online]. Available: <https://doi.org/10.13026/C2RP4B>
- [57] I. Silva et al., "Noninvasive fetal ECG: The physionet/computing in cardiology challenge 2013," in *Proc. Comput. Cardiol.*, 2013, pp. 149–152.
- [58] I. Silva et al., "Noninvasive fetal ECG: The PhysioNet/Computing in cardiology challenge 2013," 2016. [Online]. Available: <https://physionet.org/content/challenge-2013/1.0.0/>
- [59] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [60] J. Behar, J. Oster, and G. D. Clifford, "Non-invasive FECG extraction from a set of abdominal sensors," in *Proc. Comput. Cardiol.*, 2013, pp. 297–300.
- [61] F. Andreotti et al., "Non-invasive fetal ECG signal quality assessment for multichannel heart rate estimation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2793–2802, Dec. 2017.
- [62] Y. Yu et al., "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [63] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 02, pp. 107–116, 1998.
- [64] S. Santurkar et al., "How does batch normalization help optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, vol. 31, pp. 2488–2498.
- [65] I. Goodfellow et al., "Generative adversarial nets," in *Neural Information Processing Systems*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., Curran Associates, Inc., 2014. [Online]. Available: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
- [66] P. Singh and G. Pradhan, "A new ECG denoising framework using generative adversarial network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 2, pp. 759–764, Mar./Apr. 2021.
- [67] S. H. Gheshlaghi, C. N. E. Kan, and D. H. Ye, "Breast cancer histopathological image classification with adversarial image synthesis," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2021, pp. 3387–3390.
- [68] J. Kong, J. Kim, and J. Bae, "HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 17022–17033.
- [69] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. San Francisco, CA, USA: Academic, 2005, vol. 8.
- [70] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [71] J. Le Roux et al., "SDR-half-baked or well done," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 626–630.
- [72] M. Kolbæk et al., "Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1901–1913, Oct. 2017.
- [73] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 326–327, 1995.
- [74] X. Ying, "An overview of overfitting and its solutions," in *Proc. J. Phys.: Conf. Ser.*, 2019, vol. 1168, no. 2, Art. no. 022022.
- [75] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [76] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *Plos one*, vol. 16, no. 7, 2021, Art. no. e0254841.
- [77] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep CNNs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 2917–2931, Sep. 2020.
- [78] S. Mahdiani et al., "Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 5948–5951.
- [79] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization, cornell university library," 2017, *arXiv:1412.6980*.
- [80] M. R. Mohebian et al., "Single channel high noise level ECG deconvolution using optimized blind adaptive filtering and fixed-point convolution kernel compensation," *Biomed. Signal Process. Control*, vol. 57, 2020, Art. no. 101673.
- [81] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [82] H. Limaye and V. Deshmukh, "ECG noise sources and various noise removal techniques: A survey," *Int. J. Appl. Innov. Eng. Manage.*, vol. 5, no. 2, pp. 86–92, 2016.
- [83] T. Kim et al., "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2017, pp. 1857–1865.