Fetal ECG Extraction from Maternal ECG using attention-based Asymmetric CycleGAN

Mohammad Reza Mohebbian*, Seyed Shahim Vedaei, Khan A. Wahid, *Member, IEEE*, Anh Dinh, *Member, IEEE*, Hamid Reza Marateb, *Member, IEEE*

***Abstract*— Non-invasive fetal electrocardiogram (FECG) is used to monitor the electrical pulse of the fetal heart. Decomposing the FECG signal from maternal ECG (MECG) is a challenging problem due to the low amplitude of FECG, the overlap of R waves, and the potential exposure to noise from different sources. Traditional decomposition techniques, such as adaptive filters, require tuning, alignment, or pre-configuration, such as modeling the noise or desired signal. In this paper, a modified Cycle Generative Adversarial Network (CycleGAN) is introduced to map signal domains efficiently. The high correlation between maternal and fetal ECG parts decreases the performance of convolution layers. Therefore, masking attention layer which is inspired by the latent vector is implemented to improve generators. Three available datasets from the Physionet, including A&D FECG, NI-FECG and NI-FECG challenge, and one synthetic dataset using FECGSYN toolbox are used for evaluating the performance. The proposed method could map abdominal MECG to scalp FECG with an average 97.2% R-Square [CI 95%: 97.1, 97.2] and 7.8 ± 1.9 [CI 95% 6.13-9.47] Wavelet Energy based Diagnostic Distortion on A&D FECG dataset. Moreover, it achieved 99.4% [CI 95%: 97.8, 99.6], 99.3% [CI 95%: 97.5, 99.5] and 97.2% [CI 95%:93.3%, 97.1%] F1-score for QRS detection in A&D FECG, NI-FECG and NI-FECG challenge datasets, respectively. Finally, the generated synthetic dataset is used for investigating the effect of maternal and fetal heart rates on the proposed method. These results are comparable to the-state-of-the-art and has thus a potential of being a new algorithm for FECG extraction.**

***Index Terms*—Fetal ECG, CycleGAN, Blind source separation, Deep learning, CycleGAN**

I. Introduction

The electrocardiogram (ECG) signal is used as a non-invasive method of heart disorders diagnosis. Connecting electrodes on the chest, arms, hands, or legs is the traditional way of recording ECG [1]. In addition, ECG is used for fetal heart monitoring [2].

There are two ways of recording a fetal electrocardiogram (FECG). In invasive recording FECG, electrodes are attached to the fetal scalp during delivery. In this way, the risk of infection exists; however, the signal quality is excellent. In non-invasive FECG recording, maternal abdominal ECG can be utilized for FECG extraction. Emerging these extraction techniques have revolutionized fetus healthcare and enabled the clinician to continuously monitor fetal heart activities. A proper fetal heart rate involves normal mother oxygenation and transferring oxygen to the fetus. Every mechanism that induces an oxygen channel breakage may cause disturbances in the fetal heart rate.

The obtained MECG signal may be distorted with noise (baseline drift and motion artifacts) [2]. These noise sources may contribute to measurement and instrumentation failures, such as power-line disturbance, white noise, electrode connection noise, muscle contraction, electrosurgical noise, baseline wandering, and motion anomalies [3]. Slight distortion in the MECG waveform can impair the FECG extraction [4]. Therefore, using a robust decomposition strategy against noise helps in extracting and representing the fetus's heart functionality. The majority of the decomposition approaches attempt to extract FECG's QRS waves [5], [6]; however, FECG has other essential components, which can allow extensive analysis [7]. The most recognizable parts in FECG are P waves (the depolarization wave from the SA node that traverses the atria), QRS complex (ventricle depolarization), ST segments (both ventricles are depolarized completely) and T wave (ventricular repolarization) [8].

Adaptive filters are essential methods for extracting FECG components, wherein coefficients are adapted according to the signal changes in time. However, the power spectral density of the input signal affects the convergence rate of the algorithms [9]. Since the minimum mean-square error is mostly used in adaptive filter's objective function, they require a flat and uniform power spectrum to have excellent convergence. However, real-world problems include colored noise rather than white noise [10], and colored noise can drop the efficiency of adaptive filters. The least mean squares (LMS) and

Mohammad Reza Mohebbian, Seyed Shahim Vedaei, Khan A. Wahid and Anh Dinh are with Department of Electrical and Computer Engineering, University of Saskatchewan S7N 5A9, Saskatoon, Saskatchewan, Canada.

Hamid Reza Marateb is with Biomedical Engineering Department, University of Isfahan, Isfahan, Iran, and also with Department of Automatic Control, Biomedical Engineering Research Center, Universitat Politècnica de Catalunya, BarcelonaTech (UPC), Barcelona, Spain.

Corresponding author: Mohammad Reza Mohebbian (e-mail: mom158@usask.ca)

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Recursive least square (RLS), which adapted alike Weiner optimal solution are designed to work with narrowband frequencies [9]. The prerequisite of RLS and LMS is that they require a reference signal that is morphologically akin to the abdominal MECG waveform. Methods relying on temporal features, like template-based and conventional Kalman filters, are other approaches that may be listed as failing when MECG and FECG peaks overlap. The extended state Kalman filter was introduced for robust FECG extraction which could solve the QRS coincidence issue [11]. Nevertheless, they have computational complexity and unable to succeed if they could not accurately detect R-peaks.

Blind source separation strategies such as principal component analysis (PCA), independent component analysis (ICA), and periodic component analysis are substitutes of adaptive filters [12]. The primary concept of these approaches is a linear stationary mixing matrix between sources and the higher number of abdominal channels for better FECG extraction [13]. However, these methods are not adequate in insufficient Signal to Noise Ratio (SNR) circumstances and usually require specific electrode configuration and further post-processing [14].

Some researches attempted to overcome the adaptive filter or blind source separation (BSS) drawbacks using novel techniques. Mohebbian *et al.* [14] used a BSS technique for estimating the reference signal on the adaptive filter and could decompose FECG using one channel with F1-Score 96%. Zhang *et al*. [15] used singular value decomposition and smooth window and could decompose QRS of FECG with F1-Score 99%. A convolutional neural network was utilized by Zhong *et al*. [5] and could achieve 77% F1-Score for QRS wave extraction. QRS detection was also performed by Zhong *et al.* [16] using the prefix tree-based model and could achieve the F1-Score of 95%. Many methods tried to extract QRS waves; however, in decomposing FECG, other parts should also be taken into consideration. Moreover, using a general model that can be used for different subjects with different electrode displacement needs more investigation.

Simulating MECG tries to cover all artifacts in synthetic signals to enable algorithms to be tested on different scenarios, such as various noises, gestational ages, and assorted artifacts [17]. One of the famous libraries for generating synthetic MECG is fecgsyn [18], which uses realistic noise, a heart rate variability probability model, rotation maternal and fetal heart axes, fetal movement, and physiological features. The generation of FECG signals based on physiological and mathematical models needs a systematic understanding of the factors involved in producing a FECG signal. However, evolving generative adversarial approaches recently provided new perspectives for the latent vector-based data generation that learns the system's nature without prior information [19].

In this paper, Cycle Generative Adversarial Network (CycleGAN) is introduced in mapping between MECG and FECG. The idea of CycleGAN is from Hertzmann *et al.* [20], who utilized a non-parametric texture model for mapping two images. Using the weight-sharing technique for representing across domains was used by Aytar *et al*.[21] and was extended by Liu *et al.*[22] by adopting variational autoencoders and generative adversarial networks. Recently, CycleGAN has been used for non-parallel voice conversion [23], adapting image emotion [24] and style transfer in x-ray angiography [25]. Also, using deep neural networks for ECG analysis has been investigated in recent studies [26]–[28].

We have modified CycleGAN to find a map and its inversed map between maternal and fetal ECG using adversarial loss without explicit density estimation. The proposed algorithm is trained and evaluated using three datasets provided by Physionet. The remaining of the paper is arranged as follows: information about the datasets and method implementation are presented in the next section. The results of the proposed approach are described in section 3. Sections 4 and 5 include the discussion and conclusions, respectively.

II. Materials and Method

The block diagram of the proposed algorithm is depicted in Figure 1. Briefly, a sliding window is applied to the abdominal MECG and corresponding FECG signals. Then, they are normalized, and bandpass filtered and feed to the CycleGAN network for mapping and inverse mapping of MECG and FECG. Two generators are trained to receive two discriminators on adversarial training strategy. The generators are designed based on one-dimensional signal processing and should maintain the smoothness of the signal. The proposed method is evaluated in two different scenarios, including FECG signal extraction and fetal QRS detection. In FECG signal extraction, whole FECG signal components are extracted, while in fetal QRS detection, only R-wave positions are extracted. More specific details regarding each phase are given in the sub-sections below.

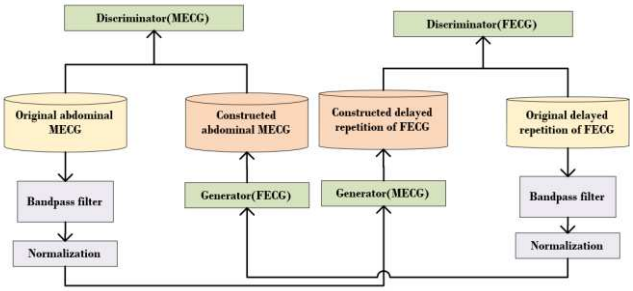

Fig. 1. The block diagram of the proposed algorithm for signal deconvolution.

2.1 Dataset

The abdominal and direct FECG (A&D FECG) [29] dataset from Physionet (https://physionet.org/content/adfecgdb/1.0.0/) was used as the main dataset that contains FECG recordings. The data contains multichannel fetal electrocardiogram (FECG) recordings obtained from a fetus scalp of 5 different women (subject 1: record r01, subject 2: record r07, subject 3: record r10, subject 4: record r04 and subject 5: record r08), between 38 and 41 weeks of gestation. Each recording consists of five minutes, four abdominal channels as well as a corresponding FECG obtained from the fetal head. All signals were sampled at 1 kHz and 16-bit resolution and bandpass filtered during acquisition (0-100 Hz) with digital filtering of the powerline. The abdominal electrode configuration consisted of four

electrodes around the belly button, a reference electrode above the symphysis of the pubic, and a common reference electrode on the left leg. These positions were constant during all recordings. Although the FECG is recorded directly from the scalp, it contaminates the maternal ECG. According to Nurani *et al.* [30], when the period is closer to delivery due to uterine contractions, the impact of maternal ECG on directly recorded FECG can also be higher.

The non-invasive FECG (NI-FECG) Physionet (https://physionet.org/content/nifecgdb/1.0.0/) was also used [31]. This dataset does not contain direct FECG signals from the scalp and only have QRS time samples for FECG. It contains 55 multichannel abdominal MECG, taken from a subject between 21 and 40 weeks of pregnancy. The electrode positions are not fixed and change sometimes to change the signal to noise ratio. The data is sampled at 1 kHz and 16-bit resolution, and bandpass filtered during acquisition (0-100 Hz). Fourteen sets, including 154, 192, 244, 274, 290, 323, 368, 444, 597, 733, 746, 811, 826, 906, are selected because other research selected almost the same sets and it facilitates the comparison [12], [14].

Furthermore, Set-A of the 2013 Physionet/Computing in Cardiology Challenge [32] is used for benchmarking (NI-FECG challenge). The archive consists of 75 abdominal ECG data recorded at a sampling rate of 1 kHz on four channels. The fetal reference R-peaks are provided for Set-A. According to many incomplete annotations [33], [34], records a33, a38, a52, a54, a71, and a74 were excluded, leaving 69 records for evaluation.

In order to train the proposed method on NI-FECG and NI-FECG challenge datasets, simulated FECG is generated based on R-R interval provided as ground truth in datasets. More details about simulation which is based on Daubechies wavelets is provided in [35] and a sample code is available at https://github.com/antecessor/FECGCycleGAN/blob/master/test/ECGSimulation.py. These simulated signals are mainly used for the training of the proposed system to be used in QRS detection and no further assessment is carried out for the extraction of FECG signals such as what is performed on the A&D FECG dataset.

The FECGSYN toolbox [36] is used, in addition to the available datasets, to model 24 maternal and fetal ECGs at various fetal and maternal heart rates, including 115, 125, 135, 145, 155, 160 bpm [37] and 65, 77, 89, 100 bpm, respectively [38].

2.2 Preprocessing

Each record of the A&D FECG dataset has a 300-second duration. Since the first and last parts of each record have more artifacts and noise, 25 seconds from the first and the last parts are truncated. Signals are downsampled five times (from 1 kHz to 200 Hz) to have less computational cost. The clinical ECG has a minimum bandwidth of 100 Hz [39] and according to Sameni and Clifford [40], the majority of the ECG relative power falls under 35 Hz, and the QRS complex has a frequency range of 10-15 Hz for FECG. Therefore, a bandpass filter with the cut-off frequencies 1 Hz and 100 Hz is applied to signals. Since the smoother signal offers better results for the proposed approach and it is not acceptable to select a narrower bandpass filter owing to losing information, the Savitzky-Golay filter [41], [42] was used in our study. The rectangular sliding window is performed on all signals; then, signals are normalized using z-score normalization [43]. The vector x_i is referred to the one abdominal MECG channel in size $1\times M$. Also, i represents the *i-th* window. Similarly, y_i represents FECG vector, for *i-th* window with size the $1\times M$, wherein M refers to the number of samples, which is set to 200 for this paper.

2.3 The Proposed Method

The goal is to map domains $X = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_1, y_2, \dots, y_P\}$, while X is $N \times M$ abdominal MECG and Y is the $P \times M$, containing the FECG signal. Like other CycleGAN approaches [23], the model contains two mappings, including $G : X \to Y$ and $F: Y \to X$. Also, two adversarial discriminators D_x and D_y should be defined, where D_x is focused to differentiate between X and $F(Y)$ and objective of D_y is to discriminate between Y and $G(X)$. Concisely, two main objectives follow adversarial loss, trying to fit the representation of the produced signal to the representation of data in the endpoint domain, and cycle consistency loss that avoid the trained mappings G and F from contradicting each other. The final objective cost is defined in Equation (1).

$$\arg\min_{G,F}\max_{D_x,D_y}\mathcal{L}(G,F,D_x,D_y)$$
$$\mathcal{L}(G,F,D_x,D_y) = \mathcal{L}_{GAN}(G,D_y,X,Y) + \mathcal{L}_{GAN}(F,D_x,Y,X) + \lambda\,\mathcal{L}_{cyc}(G,F) \quad (1)$$

Wherein,

$$\mathcal{L}_{GAN}(G,D_y,X,Y) = E_y[\log D_y(y)] + E_x[\log(1 - D_y(G(x))] \quad (2)$$

$$\mathcal{L}_{GAN}(F,D_x,Y,X) = E_x[\log D_x(x)] + E_y[\log(1 - D_x(F(y))] \quad (3)$$

$$\mathcal{L}_{cyc}(G,F) = E_x\left[\left\|F(G(x)) - x\right\|_1\right] + E_y[\left\|G(F(y)) - y\right\|_1] \quad (4)$$

where λ controls relative objectives, which is set to 4 to give more attention to cyclic consistency. Equations 2 and 3 represent the adversarial objective, which trains G and F to produce outputs identically as targets. On the other hand, objective 4 has two terms of forward and backward cycle consistency that guarantees the learned models can map individual input to a specific output. Using greater lambda reduces identity mapping.

Former GAN methods depend on the presence of one-to-one examples for training; however, the CycleGAN is capable of learning these transformations without the need to map one-to-one between the training data in source and target domains [23]. The requirement for a coupled image in the target domain is removed by having a two-step transition-first by attempting to convert it to the target domain and then back to the main image. The generator is applied to the image to map it to the target domain, then the quality of the generated image is improved by performing the generator against a discriminator. This feature enables the algorithm to train with a small dataset.

The network architecture contains two main parts, generator, and discriminator. The generator architecture is depicted in Figure 2. The purpose of the generator is to intensify waves

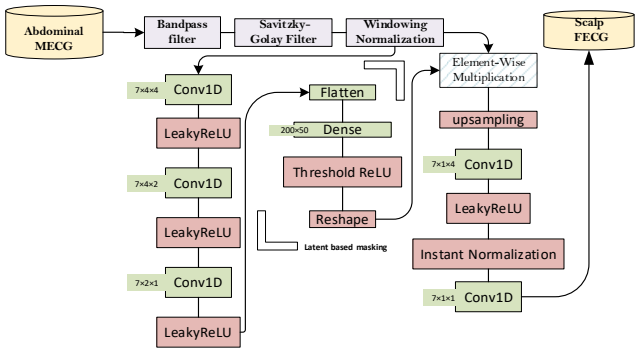

Fig. 2. The generator architecture for converting MECG to FECG.

pertaining to the target signal and to reduce the influence of waves of input signals. However, based on the analysis, we observed that the convolutional kernels, which could increase the amplitude of the waves required, could also boost the other waves. In other words, instead of suppressing the maternal R wave, it may amplify it along with enhancing the R wave of the fetus. To solve this problem, a novel masking approach is used to provide more attention to certain parts of the signal and to avoid processing parts that can increase errors. In this paper, this masking method is called latent based masking (LBM), since it works based on encoding the signal to latent vector-like autoencoder concept. The LBM increases the convergence rate and performance of the algorithm, which is elaborated in the discussion.

Briefly, a one-dimensional convolution layer (Conv1D) is applied on each channel along with the Leaky Rectified Linear Unit activation (LeakyReLU) function [44]. Three encoding layers based on Conv1D and LeakyReLU are used on all channels to get one latent vector. A dense layer is used in the next layer to compress this latent vector. One-dimensional upsampling is used to replicate latent values to create a mask with the size of the input samples. Then, threshold ReLU is used to set values below 0.5 to 0. The algorithm aims to extract features from the signal based on a cyclic objective to construct the right mask that could obtain FECG's region of interest. This mask is multiplied by the input signal to put attention on specific components. After applying LBM, the obtained signal is upsampled and passed to three Conv1D layers for extracting the desired signal. Since the Conv1D can produce high variance output (like noisy), the signal is up-sampled twice for each channel, and the Conv1D with stride two is used. This technique has a smoother output [45].

The discriminator contains three Conv1D layers with LeakyReLU that try to encode the signal to one channel and classify the input as fake or real. The generator is trained to produce signals that the discriminator identifies as actual.

The algorithm ran for 50 epochs. For the loss function, the use of l_1 and l_2 is tested based on Mean Absolute Error and Mean Square Error respectively and l_1 loss had a better performance. Using l_2 norm induces unnecessary smoothing, which causes less efficiency. Since most of the ECG signal contains low-frequency parts, l_2 norm may even converge to constant zero values. However, signal fluctuations are preserved by the standard l_1. The smoothing effect of l_2 was also observed for an electrocardiographic inverse problem for epicardial potential [46].

2.4 Validation

The performance of the proposed method is evaluated on two different datasets. In the A&D FECG dataset, the FECG signals are recorded from the fetus scalp and can be used as continuous ground truth. The leave-one-subject-out cross-validation approach was used for performance assessment. Every time, four women's signals are used for the train, and a fifth signal is used for tests. This approach is repeated five times to test all signals. In this regard, the R-squared goodness of fit [47], intra-class correlation (ICC) [48], and Bland-Altman plots or Tukey's Mean Difference Map [49] is used for reporting the performance on each subject for extracting FECG signal. The Tukey's Mean Difference Map is a statistical tool for evaluating the variations between the two measurement processes. R-squared can compare the goodness of fit of the decomposed signal, and its formulation is described in equation 5.

$$\mathrm{R}^2 = 1 - \frac{\sum(y_i - \breve{y}_i)^2}{\sum(y_i - y_{avg})^2} \qquad (5)$$

wherein, $\breve{y}_i$ is the *i-th* predicted sample and y_i is the *i-th* original sample and y_{avg} is the average of the original samples.

For diagnostic distortion analysis purposes, the mean value of the original signal and the predicted signal are subtracted from the signals. Using Daubechies 9/7 biorthogonal Wavelet filters up to 5 levels, all signals are decomposed. The QRS complex typically has the highest amplitude and the widest spectrum, therefore the QRS complex is visible on all levels. It is more noticeable in the second and third levels, though. On the first two levels, the P and T waves are not apparent and do primarily belong to four and five levels. The deviation between the original signal's Wavelet coefficients and the reconstructed signal's Wavelet coefficients is determined by the percentage root mean square difference, referred to as Wavelet PRD (WPRD). Finally, the Wavelet Energy based Diagnostic Distortion (WEDD) is calculated by the weighted average of WPRD in all levels [50].

$$WEDD = \sum_{l=1}^{L+1} w_l \; WPRD_{\mathrm{l}} = \sum_{l=1}^{L+1} w_l \sqrt{\frac{\sum_{k=1}^{K_l}\left[d_l(k) - \widetilde{d_l(k)}\right]^2}{\sum_{k=1}^{K_l}[d_l(k)]^2}}, \quad \mathrm{l} = 1,2,3,\ldots \mathrm{L} \qquad (13)$$

where, $WPRD_{\mathrm{l}}$ is the error in $l-th$ subband, and $d_l(k)$ and $\widetilde{d_l(k)}$ are the $k-th$ wavelet coefficient in $l-th$ subband of original and predicted signals, respectively. The WEDD value can be categorized in five quality groups [50], including excellent (0-4.6), very good (4.6-7), good (7-11.2), not bad (11.2-13.6) and bad (>13.6).

The QRS wave detection was performed using P&T [51] and analyzed using traditional sensitivity, positive predictive value (PPV), and F1-Score [2], [15], [16]. Equations 6, 7, and 8 explain the sensitivity, PPV, and F1-score, respectively. For calculating the performance, different time precision is used by various researches. For example, Guerrero-Martinez *et al.* [52]

used a matching window of 50 ms, while Zhang *et al*. [15] used 30 ms. In this study, every one sample in decomposed data belongs to the 5 ms due to down-sampling. Therefore, a window of 6 samples (30 ms) is used for calculating performance indices.

Following the STARD [53] and TRIPOD [54] standards, the CI 95% of the performance indices R-square, ICC, Sensitivity, PPV, and F1-score were provided, showing the reliability of the estimation.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{F1} - \text{score} = \frac{2 \times \text{PPV} \times \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}} \quad (8)$$

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (9)$$

The paired-sample t-test was used to identify whether the reconstructed signals are significantly different from the original signals especially in terms of bias. Results are reported as mean ± standard deviation, and P-values less than 0.05 were considered significant. The statistical analysis was performed using SPSS Statistics for Windows version 22 (IBM Corp. Released 2013. Armonk, NY: IBM Corp.). All algorithms are run on a system with Core-i9, 16 GB RAM, and 6 GB Graphic Cards NVIDIA GeForce GTX 1060 configurations. The proposed method was implemented in Python, and all codes are available in GitHub at https://github.com/antecessor/FECGCycleGAN.

TABLE I
THE PERFORMANCE OF THE ESTIMATED FECG SIGNAL (CI 95% ARE REPORTED IN PARENTHESIS)

Subject	1	2	3	4	5
ICC	0.98 (0.97-0.98)	0.97 (0.96-0.97)	0.94 (0.93-0.94)	0.96 (0.95-0.96)	0.97 (0.96-0.97)
R-squared	0.99 (0.99-1.00)	0.98 (0.97-0.98)	0.92 (0.91-0.92)	0.98 (0.97-0.98)	0.98 (0.97-0.98)
WEDD	7.20 % (7.05–7.35)	6.40 % (6.25-6.55)	11.20 % (11.10-11.3)	6.70 % (6.60-6.80)	7.90 % (7.76-8.04)

III. RESULTS

3.1 Results for signal extraction

Two examples of mapping MECG and FECG is depicted in Figure 3 for subject 2 and 4 for visualizing the generated MECG and extracted FECG signals.

The Tukey/Bland-Altman mean difference plot of the decomposed FECG for each subject is depicted in Figure 4. The R-Square graph is plotted in Figure 5, and on average,

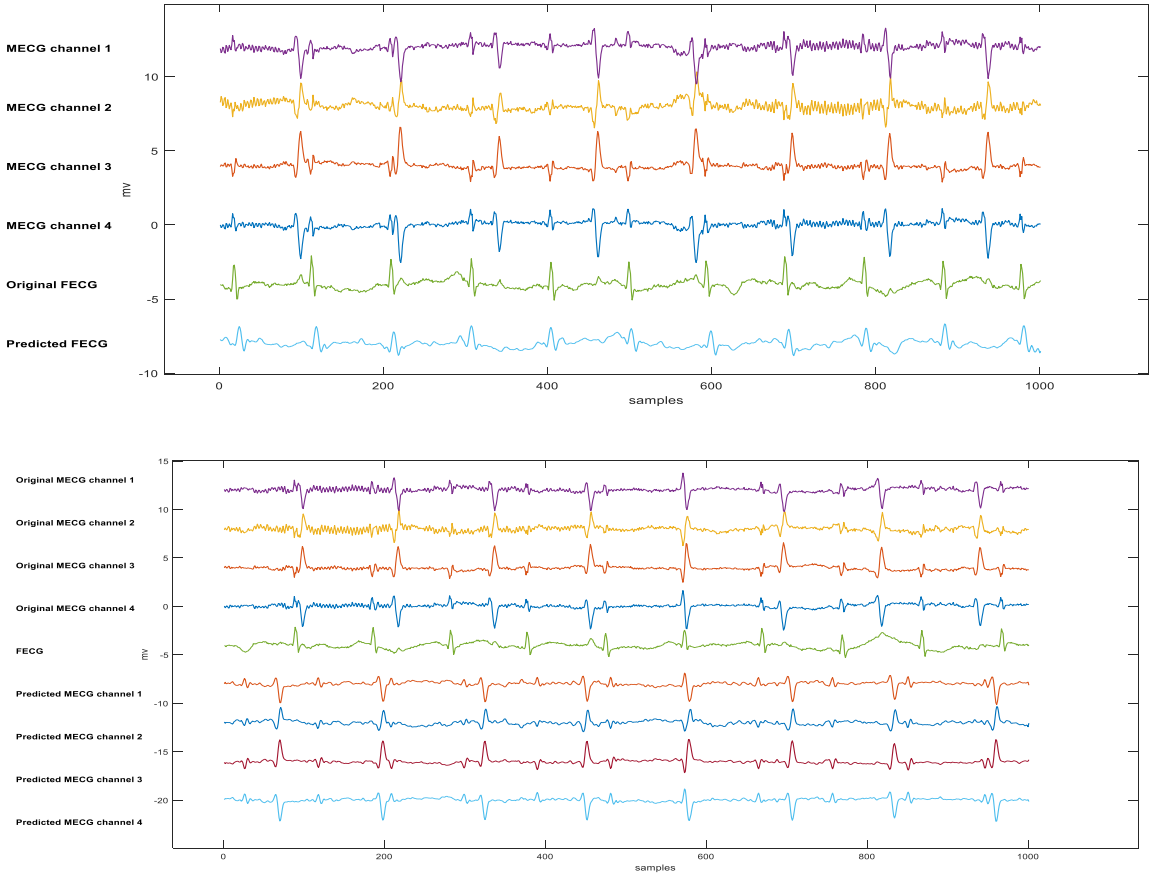

Fig 3. Two examples of FECG and MECG generation. The top one is the decomposed FECG using proposed method on 1000 samples of MECG test set for subject 4; The bottom one is the generated four channels MECG from FECG on subject 2 for 1000 samples. All signals are normalized for visualization purposes.

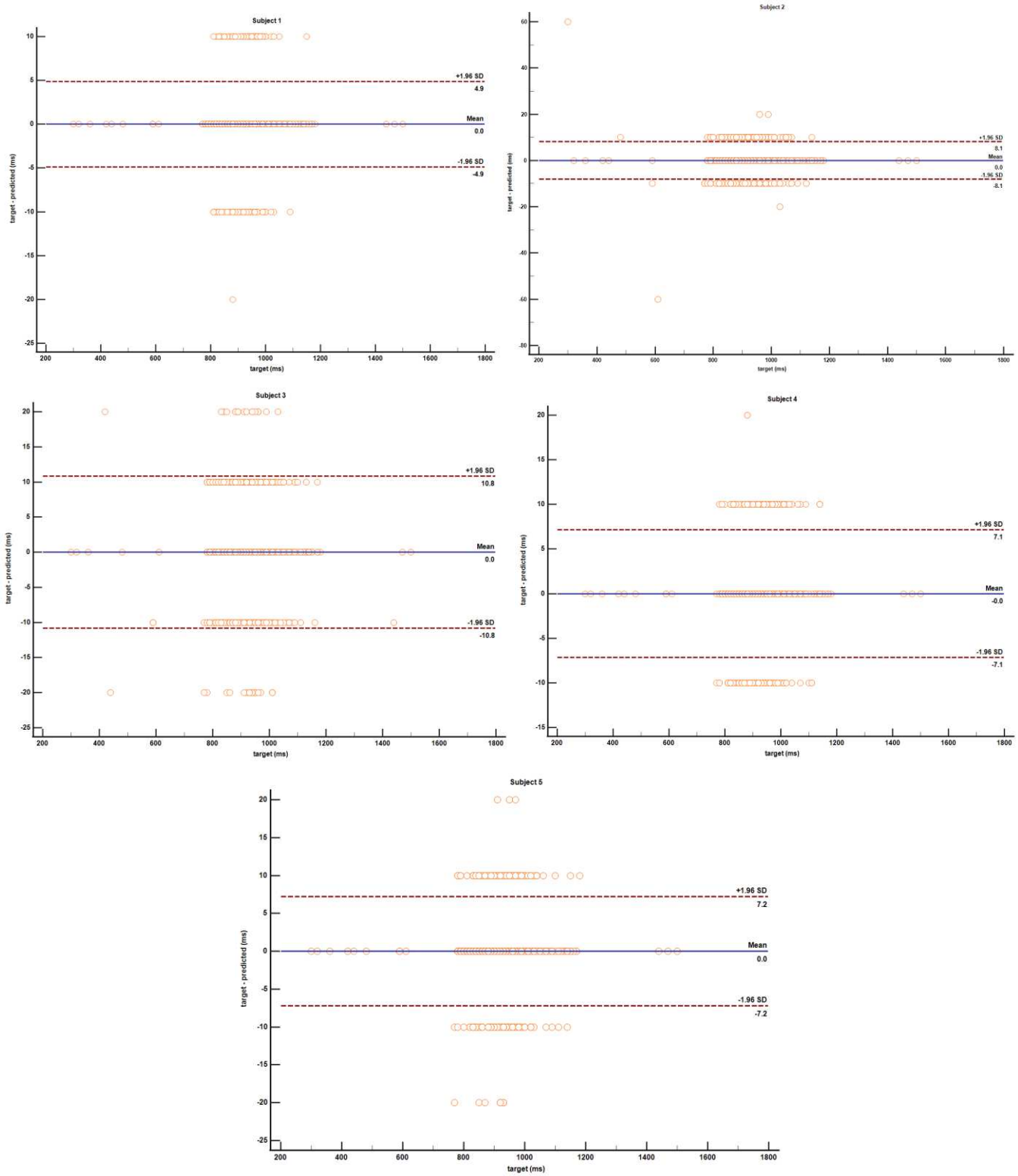

Fig. 4. The Bland-Altman plots for R-R interval of decomposed FECG signal obtained from the generator versus the R-R interval of original FECG signal for subjects 1 to 5 of A&D FECG signal. It is also demonstrated the closer 96% Bland-Altman limit values. The error of the entire R-R interval extraction, except one case in subject 2, is below the acceptance 30 ms threshold. Target: The R-R interval of original FECG; predicted: The R-R interval of predicted FECG. SD: standard deviation.

97.2% R-Squared is achieved for all subjects. The ICC, R-squared and WEDD indices are shown in Table I. All parameters suggest that the signals predicted are very similar to the original ones. According to [50], the rate of distortion is in very good and good categories for all subjects, which means that the main FECG components are retained for all subjects. Moreover, there were no significant differences between the reconstructed and original signals 1-5 (paired t-test; P-value>0.1).

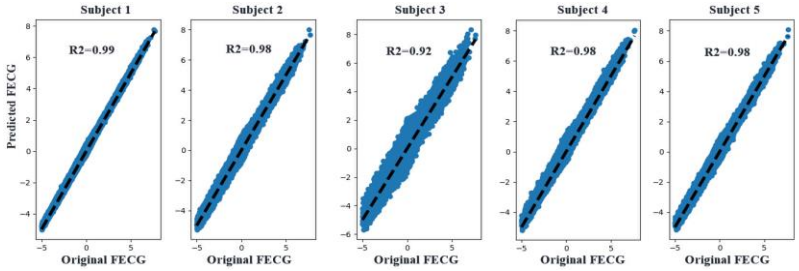

Fig. 5. The R-Squared graph for FECG signal obtained from the generator versus the original FECG signals for five subjects in the test set. The R-Square value is also shown inside subplots.

Since the training and assessment of A&D FECG dataset was based on subject-leave-out, all heart rate variabilities are not considered for the trained models. For generalization, a model is trained on all A&D FECG dataset. This model is used for testing on NI-FECG and NI-FECG challenge datasets. Moreover, this model is applied on 24 simulated MECG and FECG signals generated by FECGSYN toolbox. The boxplot for R-Squared and WEDD indices are plotted for different maternal and fetal heart rates in Figure 6. This figure shows that the performance of the system does not substantially drop if it is trained with data that cover most of the heart rate variabilities. There is no correlation between WEDD, and maternal HR (Kendall's τ_b =-0.016; P-value=0.918). However, there is a weakly negative correlation between R-square, and maternal HR but not statistically significant (Kendall's τ_b =-0.221; P-value=0.165).

3.2 Results for QRS detection

NI-FECG challenge and NI-FECG are used for QRS detection evaluation. NI-FECG is obtained from a subject in 21 and 40 weeks of pregnancy. Compared to the A&D FECG dataset, which is recorded in 38 and 41 weeks, the FECG QRS of these two datasets has less amplitude, and it should be more difficult for the algorithm to extract complete FECG. In one approach, the trained system on the A&D FECG dataset is tested on 14 signals of NI-FECG and 69 signals of NI-FECG challenge. In the second try, 4-fold cross-validation is performed on the NI-FECG dataset and NI-FECG challenge datasets, and the average and standard

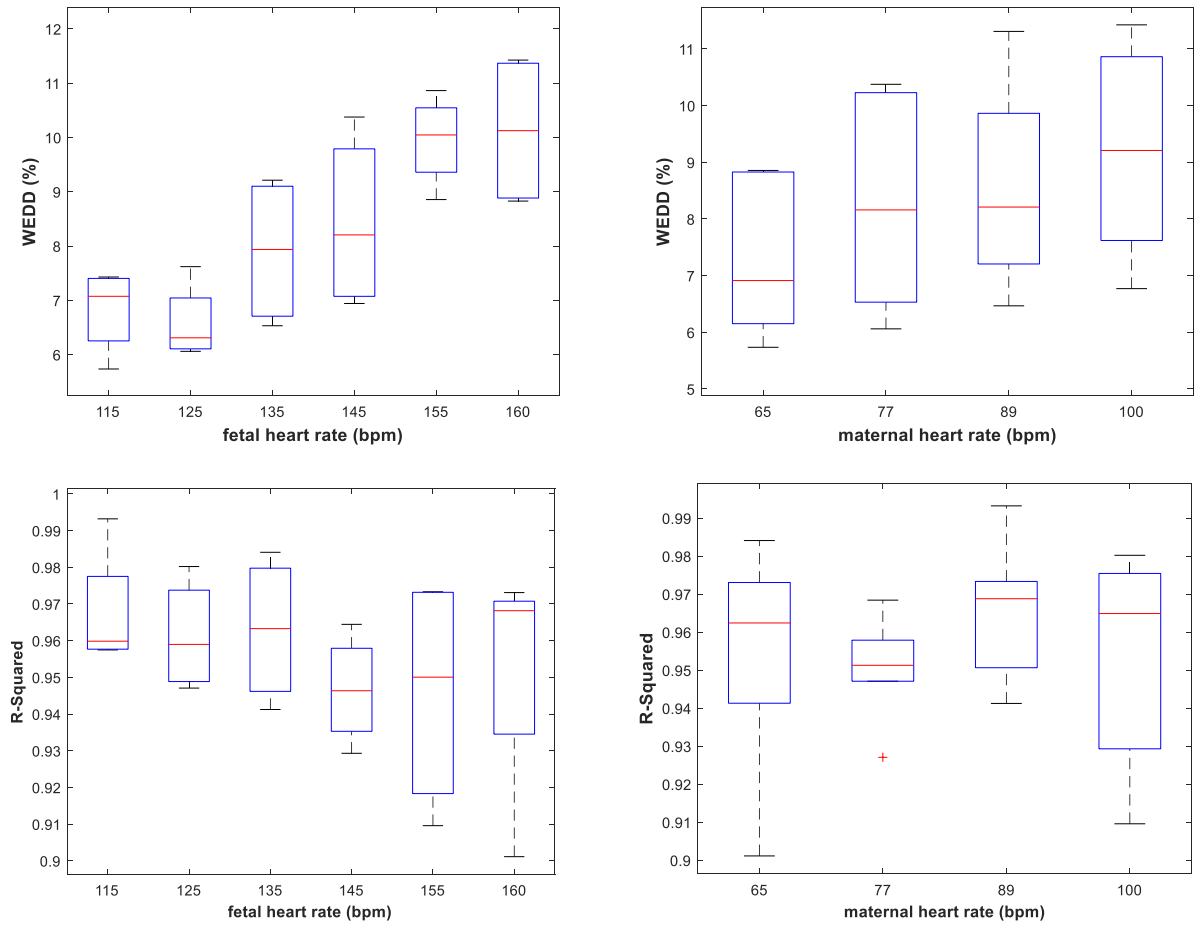

Fig. 6. Boxplots for R-Squared and WEDD indices computed on simulated FECG and MECG based on trained model using A&D FECG dataset. The trained system can extract fetal signal in different heart rates without considerable decreased performance.

deviation of test sets are reported. Since the original FECG is not provided in these two datasets, the simulated FECG based on R-R information is generated and then used for training. Table II shows the results.

IV. Discussion

4.1 The performance of the proposed method

The novelty of this algorithm was the LBM layer in deep learning architecture, which aimed to enhance efficiency. To address this argument, the LBM is excluded from the architecture, and the system is trained only based on Conv1D layers along with LeakyReLU and instance normalization. The structure of the discriminator and all other parameters remained the same. The R-Squared index for five subjects of the A&D FECG dataset is dropped to 0.86, 0.86, 0.90, 0.86 and 0.87, for subjects 1 to 5, respectively. For subject 3 (record r10), the R-Squared result is not improved sufficiently. The average maternal heart rate for this subject is higher than in others. Since the training is performed based on subject-leave-out, the cyclic cost was not trained well for different heart rate variability.

Furthermore, increasing heart rate leads to a greater chance of coincidence between fetal and maternal QRS. In cases where the fetal and maternal QRS complex occurs simultaneously or close enough, not only the fetal QRS detection Type I error increases, but also the reconstruction FECG signal may not be accurate [55]. Figure 7 shows the maternal heart rate for different subjects in the A&D FECG dataset.

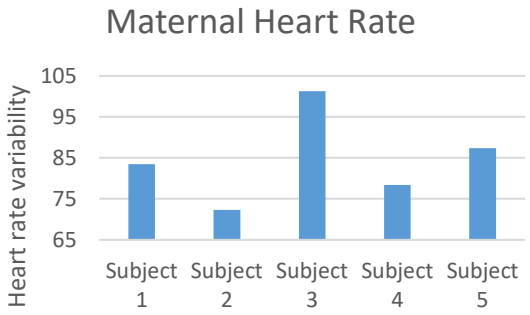

Fig. 7. The heart rate variability (HRV) of subjects in the A&D FECG dataset. Subject 3 has the highest HRV.

All algorithms include two main phases: FECG extraction and QRS detection. A fair comparison should be applied between signal extraction parts of other algorithms and proposed deep learning methods. FECG is typically influenced by the low SNR ratio, baseline drift, high amplitude, electromyogram (EMG), power-line interference, mother respiration, motion disturbances, and electrode contact loss. Nevertheless, the proposed algorithm could map MECG to FECG with high reliability according to Table 1 and Figure 3, 4, 5 and 6. For QRS detection comparison, Table 3 compares different approaches for fetal QRS complex detection.

Behar *et al.* [12] evaluated recurrent neural network (RNN), PCA, LMS and RLS, and template subtraction for QRS detection based on the P&T method [51]. They trained on 30 seconds of the NI-FECG dataset and used a 50 ms matching window. They also reported results of the trained model on a single dataset, which shows that the proposed method is robust. It is worth noting that training data within one dataset and testing on another set that is recorded using specific hardware and with different subjects at various pregnancy stages, shows whether the models are robust enough to work with similar though not identical data. The proposed method also had higher accuracy on the A&D FECG dataset over the Encoder-Decoder method [56]. They also used P&T [51] for QRS detection. The result shows that the LBM layer could help the algorithm to improve the performance. Although the proposed method achieved the same F1-score compared to SVD-SW [15], the proposed method is tested on more signals. The NI-FECG dataset had the highest accuracy when the training and test sets are from the same dataset. However, being different weeks of gestation for training and testing sets, result in different quality and different intensity of fetal cardiac activity which can be stated as the critical factor for dropped accuracy in our experiments.

The average performance of the proposed method on the NI-FECG challenge dataset is less when it is performed on other datasets. In some noisy abdominal records with small FECG amplitude, like a29, a38, a40, a42, a52, a53, a54, a56, a61, and a65, the performance considerably deteriorated which affect the overall performance. Moreover, the trained model on A&D FECG, which evaluated on NI-FECG challenge dataset, had less performance than other state-of-the-art models. This could be because the fetal R waves are sharper in A&D FECG dataset than R waves in the NI-FECG dataset. Difficulties may arise when the fetal ECG R waves are measured through the thick abdominal wall, or there are other isolating layers (e.g., vernix caseosa) or merely since the pregnancy is in the initial days of the third trimester.

Nevertheless, the proposed method outperformed to Behar *et al*. [34] that used a combination of template subtraction and ICA. Varanini *et al.* [33] method outperformed the proposed QRS detection algorithm. They proposed a signal extraction technique based on ICA and a post-processing method specialized for detecting the QRS on the NI-FECG challenge dataset. Their QRS detection algorithm is performed on extracted FECG signals using the proposed method, and the

TABLE II
PERFORMANCE OF THE QRS DETECTION BASED ON P&T METHOD; MEAN ± STD (CI 95% IS REPORTED IN PARENTHESIS)

Train	Test	F1-score (%)	PPV (%)	Sensitivity (%)	# tested signals
A&D FECG	A&D FECG	99.4 ± 0.4 (97.8-99.6)	99.3 ± 0.4 (97.6 - 99.4)	99.3 ± 0.6 (97.8- 99.5)	5
A&D FECG	NI-FECG	96.4 (94.5-97.0)	97.5 (94.9-97.9)	95.3 (94.1-97.0)	14
NI-FECG	NI-FECG	99.3 ± 1.6 (97.5-99.5)	99.5 ± 0.9 (97.6-99.5)	99.1± 1.1 (97.4-99.4)	14
A&D FECG	NI-FECG challenge	91.3 (89.8-93.1)	92.3 (90.2-93.5)	90.5 (88.1-90.9)	69
NI-FECG challenge	NI-FECG challenge	96.3 ± 0.3 (93.3-97.1)	96.7 ± 0.4 (94.6-97.1)	95.9 ± 0.6 (94.6-96.9)	69

TABLE III
COMPARING THE RESULT OF THE DIFFERENT METHODS FOR DECOMPOSING FETAL ECG

Method	F1-score (%)	Dataset	QRS detection method	Number of channels	Matching window length (ms)	Number of tested signals
TS and ES-RNN [12]	97.2	NI-FECG	P&T	1 Abdominal	50	14
	90.2	Train on NI-FECG and test on private dataset				
TS and PCA [12]	95.4	NI-FECG		1 Abdominal		14
	89.3	Train on NI-FECG and test on private dataset				
LMS [12]	95.4	NI-FECG		1 Abdominal		14
	87.9	Train on NI-FECG and test on private dataset				
RLS [12]	95.9	NI-FECG		1 Abdominal		14
	88.2	Train on NI-FECG and test on private dataset				
OBACKC [14]	95.6	NI-FECG	Thresholding	1 Thoracic		14
SVD-SW [15]	99.4	A&D FECG	P&T	1 Abdominal		2
Encoder-Decoder [48]	94.1	A&D FECG		1 Abdominal		5
Behar *et al*. [34]	95.9	NI-FECG challenge		4 Abdominal		69
Varanini *et al.* [33]	99.0	NI-FECG challenge	Customized	4 Abdominal		69
Proposed Method	96.4	Train on A&D FECG and test on NI-FECG	P&T	4 Abdominal	30	14
	99.3	NI-FECG				14
	99.4	A&D FECG				5
	91.3	Train on A&D FECG and test on NI-FECG challenge				69
	96.3	NI-FECG challenge				69
	90.1	Train on A&D FECG and test on NI-FECG		1 Abdominal		14
	95.5	NI-FECG				14
	95.9	A&D FECG				5
	88.7	Train on A&D FECG and test on NI-FECG challenge				69
	94.8	NI-FECG challenge				69
	97.2	NI-FECG challenge	Customized			69

TS: Template subtraction, ES-RNN: echo state for the recurrent neural network, PCA: principal component analysis, LMS: Least Mean Square, RLS: Recursive Least Square, ICA: independent component analysis; OBACKC: optimized blind adaptive filtering using convolution kernel compensation; SVD-SW: singular value decomposition and smooth windowing.

performance increased to an average of 97.2% using 4-fold cross-validation. Warmerdam *et al.* [57] used a multichannel hierarchical probabilistic system, which incorporates ECG waveform and heart rate predictive models to detect fetal R peaks in NI-FECG challenge dataset. They reported 99.6% accuracy defined as TP/(TP+FP+FN) which outperform Varanini *et al* [33] method with the accuracy of 98.6%. The accuracy of our proposed method was 98.2% on QRS detection. Nevertheless, Warmerdam *et al.* did not provide sensitivity or F1-Score for better comparison.

4.2 Sensitivity analysis

The variation of λ_2 coefficient, which is in the objective function of CycleGAN, is assessed by adjusting from 0.1 to 40, and the sensitivity analysis plot is depicted in Figure 8. Concisely, the cyclic and identity of the system could be preserved in 2 < Lambda<10. The higher the Lambda meaning, the less precision in FECG QRS detection owing to inexact FECG QRS boosting. In comparison, using smaller Lambda does not allow the algorithm to converge owing to more concentration on identity.

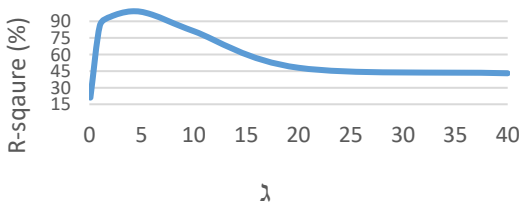

Fig. 8. Sensitivity analysis of Lambda for subject 1 in the A&D FECG dataset.

Other network parameters are also analyzed by utilizing sensitivity analysis. In one step, kernel size of the entire layers is changed from three to six. However, no difference in the performance and training time is found. Then, activation functions such as LeakyReLU, Tanh and Sigmoid are tested. The proposed method could only converge with LeakyReLU activation function. Finally, the number of Conv1D layers in LBM is decreased from 3 to 2 and the performance did not change but the training time decreased about 56 ± 16 seconds in average. Thus, the proposed method is stable against changing network parameters, although it is sensitive to initial random numbers for training which can be eliminated by assigning appropriate seed or better random generator distribution.

4.3 Effect of channel reduction, signal to noise ratio and bandwidth on performance

Optimizing the number of channels can decrease redundant information, decrease noise, and make the recording more pleasant for clinicians and mothers. According to Figure 9, reducing the channel can decrease R-Square. Maternal R-wave is always visible in all MECG channels; however, fetal R-wave is not always recognizable in MECG channels. Thus, choosing an appropriate channel, for example, with higher SNR, can improve the performance.

One can ask that since decreasing the number of channels may affect the performance on R-squared, the LBM layer is not effective as the number of channels. As a justification, in Figure 9, the R-square with and without the LBM layer is calculated. It shows that LBM could increase the performance even for one channel. In fact, The R-square values with LBM layer were

significantly better than those without the LBM layer (Wilcoxon Signed Rank Test; P-value<0.001). On the other hand, although, R-Squared is decreased significantly for one channel, the QRS detection performance does not decrease substantially. In this regard, the F1-score of the QRS detection for NI-FECG and A&D FECG datasets are decreased from 99.3% and 99.4% to 95.5% and 95.9%, respectively. It is worth noting that utilizing only one channel results in a bit noisy reconstructed FECG signal. However, the P&T approach for QRS detection was able to recover the QRS complex in such noisy circumstances. The QRS detection accuracy depends on the proposed decomposition method and QRS detection algorithm; therefore, comparing the performance of the QRS detection does not clearly explain the effectiveness of the proposed method.

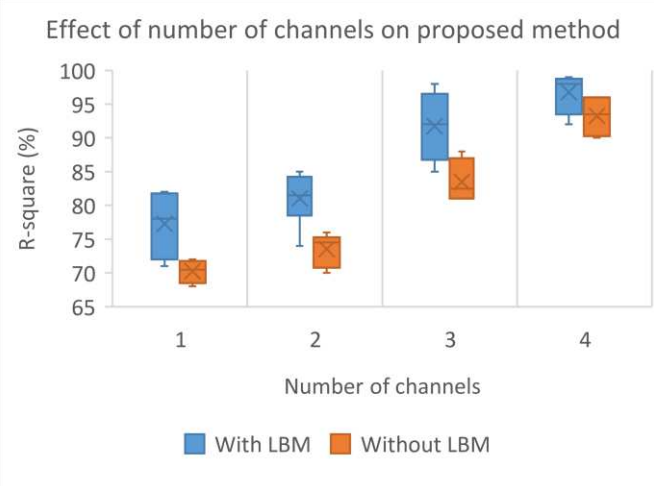

Fig. 9. The effect of number of MECG channels on R-Squared acquired for FECG signal extraction on test NI-FECG dataset with and without LBM layer. Using attention mechanism could improve the performance irrespective to the utilized number of channels.

As discussed earlier, the SNR is one of the important factors which should be considered for selecting the channel. The effect of noise on the proposed method is evaluated with three different scenarios: First, an EMG noise with different amplitudes is added to the entire channels of multichannel MECG signal as the input of the proposed method. Second, four-channel MECG is used again as the input of the proposed method, but the same noise is added to only one channel. Finally, the same noise is added to one channel, and only one channel is used as the input of the proposed method. Figure 10 depicts the SNR of signals, and the R-square acquired for subject 1 in the A&D FECG dataset based on training on other subjects with noisy signals. It shows that the multichannel approach is more robust against noise. According to the generator architecture in Figure 2, a convolutional layer with one filter is used to combine channel features. This layer can be defined as a linear projection of the feature maps stacks. If the network is trained with noisy data and only one channel is noisy, this linear projection can decide to decrease the weights of a noisy channel.

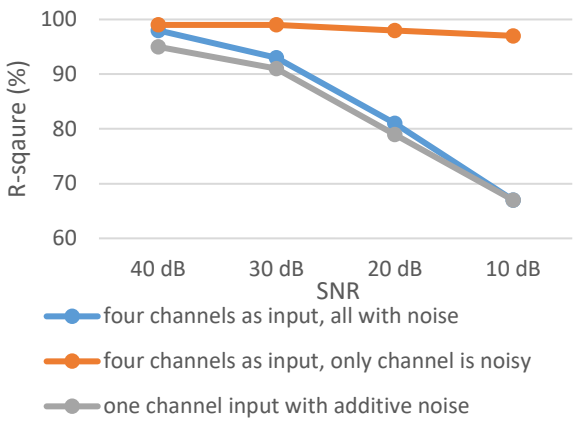

Fig. 10. The relation between noise and R-square for subject 1 in the A&D FECG dataset for different scenarios. The multichannel approach is more stable against noise.

4.4 Limitations, merits and future works

Although in QRS detection, the approach is not computationally efficient and there are works in literature that can still outperform the approach suggested, the flowing issues could be considered:

First, the detection of QRS is not the primary task that can be supported by the suggested method. In order of words, the proposed method can extract the complete FECG signal, and a QRS detection algorithm is added on top of it to demonstrate one of the capabilities of the proposed method. The WEDD index, which is used for showing how much distortion exist between predicted FECG and original FECG, can show that the proposed method can maintain important waveforms (Table I and Figure 6). Second, the proposed system trains on unpaired samples. Therefore, the proposed system allows two separate datasets which only have MECG and FECG to be trained for FECG extraction. This mechanism can be used to eliminate data availability limitations. Finally, the trained model can process ECG signals and analyze significant waves such as P, Q, R, S and T. This capability can stimulate the idea that CycleGAN can also be used to map other signals, such as ECG to PPG or vice versa [58].

One of the limitations of this study is the complexity of the model, which is common in deep learning algorithms. This complexity may increase the computational cost and maybe a burden for embedded systems, which can be eliminated by advances in hardware technology and using GPUs with less power consumption. The total training time is 10400 seconds, representing 208 seconds for each epoch, whereas the processing of 1000 four-channel signal samples requires a total of 0.14 seconds. Also, larger sample size is added to show whether this approach can operate efficiently in all circumstances.

V. Conclusion

A novel architecture for cycle generative adversarial neural network is investigated for mapping maternal and fetal ECG. The proposed method is applied to two datasets for evaluating the fetal ECG decomposition. The abdominal and direct FECG

(A&D FECG) used for fetal ECG from four maternal abdominal ECGs and could achieve an average 97.2% R-Square [CI 95%: 97.1%, 97.2%] as the goodness of fit and 99.4 % F1-score [CI 95%: 97.8%, 99.6%] for QRS estimation based on subject-leave-out validation. On the other hand, the non-invasive FECG (NI-FECG) and NI-FECG challenge datasets are also used for proving the robustness of fetal QRS estimation and could achieve 99.3% F1-score [CI 95%: 97.5%, 99.5%] and 97.2% F1-score [CI 95%:93.3%, 97.1%], respectively. A synthetic dataset is generated for investigating the effect of maternal and fetal heart rates on the performance which showed that the proposed method can be used in various fetal and maternal heart rate variations. Such results are comparable to the-state-of-the-art and therefore the proposed method is promising for FECG extraction.

References

[1] P. R. Jeffries, S. Woolf, and B. Linde, "Technology-based vs. traditional instruction: A comparison of two methodsfor teaching the skill of performing a 12-lead ecg," *Nurs. Educ. Perspect.*, vol. 24, no. 2, pp. 70–74, 2003.

[2] M. A. Hasan, M. I. Ibrahimy, and M. B. I. Reaz, "Techniques of FECG signal analysis: detection and processing for fetal monitoring," *WIT Trans. Biomed. Health*, vol. 12, pp. 295–305, 2007.

[3] G. D. Clifford, F. Azuaje, and P. Mcsharry, "ECG statistics, noise, artifacts, and missing data," *Adv. Methods Tools ECG Data Anal.*, vol. 6, p. 18, 2006.

[4] E. R. Ferrara and B. Widraw, "Fetal electrocardiogram enhancement by time-sequenced adaptive filtering," *IEEE Trans. Biomed. Eng.*, no. 6, pp. 458–460, 1982.

[5] W. Zhong, L. Liao, X. Guo, and G. Wang, "A deep learning approach for fetal QRS complex detection," *Physiol. Meas.*, vol. 39, no. 4, p. 045004, 2018.

[6] M. Varanini, G. Tartarisco, R. Balocchi, A. Macerata, G. Pioggia, and L. Billeci, "A new method for QRS complex detection in multichannel ECG: Application to self-monitoring of fetal health," *Comput. Biol. Med.*, vol. 85, pp. 125–134, 2017.

[7] R. A. Shepoval'nikov, A. P. Nemirko, A. N. Kalinichenko, and V. V. Abramchenko, "Investigation of time, amplitude, and frequency parameters of a direct fetal ECG signal during labor and delivery," *Pattern Recognit. Image Anal.*, vol. 16, no. 1, pp. 74–76, 2006.

[8] G. D. Clifford, I. Silva, J. Behar, and G. B. Moody, "Non-invasive fetal ECG analysis," *Physiol. Meas.*, vol. 35, no. 8, p. 1521, 2014.

[9] B. Rafaely and S. J. Elliot, "A computationally efficient frequency-domain LMS algorithm with constraints on the adaptive filter," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1649–1655, 2000.

[10] D. Mumford and A. Desolneux, *Pattern theory: the stochastic analysis of real-world signals*. CRC Press, 2010.

[11] 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, 2013.

[12] J. Behar, A. Johnson, G. D. Clifford, and J. Oster, "A comparison of single channel fetal ECG extraction methods," *Ann. Biomed. Eng.*, vol. 42, no. 6, pp. 1340–1353, 2014.

[13] R. Martinek *et al.*, "Comparative effectiveness of ICA and PCA in extraction of fetal ECG from abdominal signals: Toward non-invasive fetal monitoring," *Front. Physiol.*, vol. 9, p. 648, 2018.

[14] M. R. Mohebbian, M. W. Alam, K. A. Wahid, and A. Dinh, "Single channel high noise level ECG deconvolution using optimized blind adaptive filtering and fixed-point convolution kernel compensation," *Biomed. Signal Process. Control*, vol. 57, p. 101673, 2020.

[15] N. Zhang *et al.*, "A novel technique for fetal ECG extraction using single-channel abdominal recording," *Sensors*, vol. 17, no. 3, p. 457, 2017.

[16] W. Zhong, X. Guo, and G. Wang, "QRStree: A prefix tree-based model to fetal QRS complexes detection," *PloS One*, vol. 14, no. 10, 2019.

[17] A. Josko and R. J. Rak, "Effective simulation of signals for testing ECG analyzer," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 3, pp. 1019–1024, 2005.

[18] J. Behar, F. Andreotti, S. Zaunseder, Q. Li, J. Oster, and G. D. Clifford, "An ECG simulator for generating maternal-foetal activity mixtures on abdominal ECG recordings," *Physiol. Meas.*, vol. 35, no. 8, p. 1537, 2014.

[19] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[20] L. Ying, A. Hertzmann, H. Biermann, and D. Zorin, "Texture and shape synthesis on surfaces," in *Rendering Techniques 2001*, Springer, 2001, pp. 301–312.

[21] Y. Aytar, L. Castrejon, C. Vondrick, H. Pirsiavash, and A. Torralba, "Cross-modal scene networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2303–2314, 2017.

[22] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Advances in neural information processing systems*, 2016, pp. 469–477.

[23] T. Kaneko, H. Kameoka, K. Tanaka, and N. Hojo, "Cyclegan-vc2: Improved cyclegan-based non-parallel voice conversion," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 6820–6824.

[24] S. Zhao *et al.*, "CycleEmotionGAN: Emotional semantic consistency preserved cycleGAN for adapting image emotions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 2620–2627.

[25] O. Tmenova, R. Martin, and L. Duong, "CycleGAN for style transfer in X-ray angiography," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 10, pp. 1785–1794, 2019.

[26] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1232–1240, 2019.

[27] B. Taji, A. D. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124–1131, 2017.

[28] T. H. Linh, S. Osowski, and M. Stodolski, "On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1224–1231, 2003.

[29] J. Jezewski, A. Matonia, T. Kupka, D. Roj, and R. Czabanski, "Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram," *Biomed. Tech. Eng.*, vol. 57, no. 5, pp. 383–394, 2012.

[30] R. Nurani, E. Chandraharan, V. Lowe, A. Ugwumadu, and S. Arulkumaran, "Misidentification of maternal heart rate as fetal on cardiotocography during the second stage of labor: the role of the fetal electrocardiograph," *Acta Obstet. Gynecol. Scand.*, vol. 91, no. 12, pp. 1428–1432, 2012.

[31] A. L. Goldberger *et al.*, "Components of a new research resource for complex physiologic signals," *PhysioBank PhysioToolkit Physionet*, 2000.

[32] I. Silva *et al.*, "Noninvasive fetal ECG: the PhysioNet/computing in cardiology challenge 2013," in *Computing in Cardiology 2013*, 2013, pp. 149–152.

[33] M. Varanini, G. Tartarisco, L. Billeci, A. Macerata, G. Pioggia, and R. Balocchi, "An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG," *Physiol. Meas.*, vol. 35, no. 8, p. 1607, 2014.

[34] 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, p. 1569, 2014.

[35] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, 2003.

[36] E. C. G. An, "model for simulating maternal-foetal activity mixtures on abdominal ECG recordings/Behar J., Andreotti F., Zaunseder S. et al," *Physiol Meas*, vol. 35, no. 8, pp. 1537–1550, 2014.

[37] S. P. Von Steinburg *et al.*, "What is the 'normal' fetal heart rate?," *PeerJ*, vol. 1, p. e82, 2013.

[38] J. Patrick, K. Campbell, L. Carmichael, R. Natale, and B. Richardson, "Daily relationships between fetal and maternal heart rates at 38 to 40 weeks of pregnancy.," *Can. Med. Assoc. J.*, vol. 124, no. 9, p. 1177, 1981.

[39] J. J. Bailey *et al.*, "Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association.," *Circulation*, vol. 81, no. 2, pp. 730–739, 1990.

[40] R. Sameni and G. D. Clifford, "A review of fetal ECG signal processing; issues and promising directions," *Open Pacing Electrophysiol. Ther. J.*, vol. 3, p. 4, 2010.

[41] S. Das and M. Chakraborty, "QRS detection algorithm using Savitzky-Golay filter," *ACEEE Int J Signal Image Process.*, vol. 3, no. 01, pp. 55–58, 2012.

[42] N. Rastogi and R. Mehra, "Analysis of Savitzky-Golay filter for baseline wander cancellation in ECG using wavelets," *Int J Eng Sci Emerg Technol*, vol. 6, no. 1, pp. 15–23, 2013.

[43] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *ArXiv Prepr. ArXiv150306462*, 2015.

[44] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *ArXiv Prepr. ArXiv150500853*, 2015.

[45] X. Fang, M. Wang, A. Shamir, and S.-M. Hu, "Learning Explicit Smoothing Kernels for Joint Image Filtering," in *Computer Graphics Forum*, 2019, vol. 38, no. 7, pp. 181–190.

[46] S. Ghosh and Y. Rudy, "Application of l1-norm regularization to epicardial potential solution of the inverse electrocardiography problem," *Ann. Biomed. Eng.*, vol. 37, no. 5, pp. 902–912, 2009.

[47] J. Miles, "R squared, adjusted R squared," *Wiley StatsRef Stat. Ref. Online*, 2014.

[48] S. Mehta *et al.*, "Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies," *Stat. Med.*, vol. 37, no. 18, pp. 2734–2752, 2018.

[49] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Int. J. Nurs. Stud.*, vol. 47, no. 8, pp. 931–936, Aug. 2010, doi: 10.1016/j.ijnurstu.2009.10.001.

[50] M. S. Manikandan and S. Dandapat, "Wavelet energy based diagnostic distortion measure for ECG," *Biomed. Signal Process. Control*, vol. 2, no. 2, pp. 80–96, 2007.

[51] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, no. 3, pp. 230–236, 1985.

[52] J. F. Guerrero-Martinez, M. Martinez-Sober, M. Bataller-Mompean, and J. R. Magdalena-Benedito, "New algorithm for fetal QRS detection in surface abdominal records," in *2006 Computers in Cardiology*, 2006, pp. 441–444.

[53] P. M. Bossuyt *et al.*, "STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies," *Clin. Chem.*, vol. 61, no. 12, pp. 1446–1452, 2015.

[54] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) The TRIPOD Statement," *Circulation*, vol. 131, no. 2, pp. 211–219, 2015.

[55] A. Matonia, J. Jezewski, T. Kupka, K. Horoba, J. Wrobel, and A. Gacek, "The influence of coincidence of fetal and maternal QRS complexes on fetal heart rate reliability," *Med. Biol. Eng. Comput.*, vol. 44, no. 5, pp. 393–403, 2006.

[56] W. Zhong, L. Liao, X. Guo, and G. Wang, “Fetal electrocardiography extraction with residual convolutional encoder–decoder networks,” *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 4, pp. 1081–1089, 2019.

[57] G. J. Warmerdam, R. Vullings, L. Schmitt, J. O. Van Laar, and J. W. Bergmans, “Hierarchical probabilistic framework for fetal R-peak detection, using ECG waveform and heart rate information,” *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4388–4397, 2018.

[58] P. Sarkar and A. Etemad, “CardioGAN: Attentive Generative Adversarial Network with Dual Discriminators for Synthesis of ECG from PPG,” *ArXiv Prepr. ArXiv201000104*, 2020.