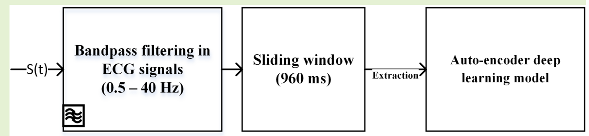
A Dual Attention-based Auto-encoder Model For Fetal ECG Extraction From Abdominal Signals

Hamidreza Ghonchi, and Vahid Abolghasemi, *Senior Member, IEEE*

***Abstract*— Fetal electrocardiogram (FECG) signals contain important information about the conditions of the fetus during pregnancy. Currently, pure FECG signals can only be obtained through an invasive acquisition process which is life-threatening to both mother and fetus. In this study, single-channel ECG signals from the mother's abdomen are analysed with the aim of extracting the clean FECG waveform. This is a challenging task due to the very low amplitude of the FECG, various noises involved in the signal acquisition, and the overlap of R waves. To address this problem, we propose a novel convolutional auto-encoder network architecture to learn and extract the FECG patterns. The proposed model is equipped with a dual attention model, composed of squeeze-and-excitation and channel-wise modules, in the encoder and decoder blocks, respectively. It also benefits from a bidirectional long short-term memory (LSTM) layer. This combination allows the proposed network to accurately attend to and extract FECG signals from abdominal data. Three well-established datasets are considered in our experiments. The obtained results of FECG extraction are promising and confirm the effectiveness of using attention modules within the deep learning model. The results also suggest that the proposed auto-encoder network can accurately extract the fetal ECG signals where no information about maternal ECG is available.**



***Index Terms*— Electrocardiography, deep learning, Fetal ECG, Attention layer**

I. Introduction

WORLDWIDE, around 2.65 million babies are stillborn each year, and heart problems are the leading cause of death due to birth defects. According to the American Heart Association, in the United States, about 40,000 children are born with heart disease each year, and at least eight out of every 1,000 babies have a heart defect [1]. As a result, there is a need for methods to prevent and reduce fetal mortality during pregnancy. Many congenital heart defects, such as fetal arrhythmia, which is mainly associated with intrauterine deterioration, are the primary reasons for fetal demise [2]. These defects should be detected before birth and during pregnancy. Hence, monitoring the fetal heart is clinically important for assessing the conditions of the fetus.

Non-invasive FECG, performed by placing electrodes on the mother's abdomen, is a promising option for standard fetal heart monitoring. This technology has significantly advanced in recent years and is safe for the fetus and easier to be used by staff [3]. However, it cannot provide a pure FECG but a mixture of FECG, maternal ECG (MECG), and noise. Hence, further processing is required to extract FECG from the recording. The extraction process faces several major challenges, such as the very low amplitude of the FECG (compared to MECG), the overlap of R waves, and various noises including maternal muscle movement, respiratory activity, fetal movement, etc. Consequently, conventional filtering techniques cannot fully extract the FECG but can be regarded as a pre-processing stage to remove some adverse disturbances that do not overlap with FECG [4].

Abdominal ECG (AECG), collected from the mother's abdomen, is a mixture of MECG, FECG, and noise. Various methods have been proposed to separate these signals and find clean FECG. Blind source separation (BSS), template subtraction (TS), and filtering techniques are common methods used to extract FECG. BSS methods assume that the abdominal signal is a combination of independent sources, including FECG, MECG, and noise [5]. As a result, by finding the signal source matrix, BSS estimates the FECG signal, which is computationally very expensive. Other traditional models like independent component analysis (ICA) [6], [7], and non-negative matrix factorization (NMF) [8] have also been studied. However, these methods have convergence issues and are costly in terms of memory storage and resources, and are not very accurate. In contrast, TS methods are inexpensive with less computational cost [9]. TS can make an adaptive pattern for MECG which is then subtracted from the entire AECG signal to eventually extract FECG. However, these methods require knowledge about the exact maternal QRS (mQRS) location in order to make a related MECG pattern and extract the FECG signal. Recently, deep learning-based approaches have been reported to be effective in this field. These methods have considerably improved the accuracy of fetal QRS (fQRS) detection compared to traditional ones [10]–[13]. However, few studies showed that deep learning can also directly extract FECG signals from AECG when no knowledge about MECG is available. These studies have shown that FECG can be extracted from AECG by following specific patterns (treated as noise) in the input signal [14]–[16]. Most related studies on the extraction of fetal ECG require access to the associated maternal ECG (or at least the exact location of maternal QRS). This is a limiting factor as maternal ECG is not often available and thus QRS detection would be a

H. Ghonchi was with the Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

V. Abolghasemi is with the school of computer science and electronic engineering, University of Essex, Colchester, UK. (email: v.abolghasemi@essex.ac.uk)

Manuscript received xxxx xx, 2021; revised xxxx xx, 2021.

complex and erroneous task. This can significantly reduce the performance of FECG extraction.

In this paper, we address FECG extraction from single-channel AECG using a novel deep learning model. The proposed model benefits from two attention modules and a bidirectional LSTM layer in order to keep all past and future information from signals and find (and extract) the best pattern of FECG signal from the AECG signal. We first apply some filters on the raw AECG signals as a pre-processing stage in order to make the signal smoother and remove unwanted noises. Then, the data is subsampled to 200 Hz to reduce the computational cost and to provide a balanced and consistent platform among both datasets. An automatic windowing procedure is also performed to generate the labelled data for training. According to the literature, [14], [15], the attention modules can help deep learning models to find and extract specific signal patterns. Therefore, the proposed auto-encoder model in this study is equipped with novel attention modules to focus on the extraction of FECG patterns. The proposed architecture is further modified so that the extracted FECGs are very similar to the original signals. In summary, the major contributions of this work are as follows:

- A novel approach to extract fetal ECG from abdominal signals without the need for prior knowledge on maternal ECG.
- A well-designed dual-attention module within the proposed deep architecture to capture fetal ECG features.
- An innovative bi-directional LSTM layer with the aim of preserving the temporal properties of the data during the signal separation.
- A generalisable model with high signal extraction accuracy verified by three independent real datasets.
- A relatively low computational complexity during training and testing.

The rest of the paper is organised as follows. Section II is devoted to reviewing the related works. In Section III, details of the proposed approach including datasets, preprocessing, and deep learning model are provided. Section IV demonstrates the results of our experiments. Finally, the concluding remarks are given in section V.

II. Related works

Recently, Zhong et al. [10] proposed a method using deep learning to detect complex fetal QRS without cancelling MECG signals. A three-layer one-dimensional convolutional neural network (1D-CNN) was used as a classifier in this paper. They achieved 77.38% accuracy for detecting complex fetal QRS. In another work, Lee et al. [17] developed a deep network based on a two-dimensional CNN for detecting fetal ECG peak in AECG signal on four-channel raw ECG signals. Their model is deeper than Zhong's model [10] where 7 convolutional layers were used to detect fetal QRS. Vo et al. [18] reported a model using deep learning to detect fetal QRS complexes with four-channel ECG signals. Their deep learning model is based on ResNet and one-dimensional octave convolution to reduce memory and computational cost. They achieved 91.1% accuracy and saved more than 50% computational cost. Huque et al. [19] used a hidden Markov model (HMM) to detect fetal QRS from ECG signals. They used ICA to separate maternal ECG and cancel the MECG signal in order to extract the FECG signal. Then, they trained HMM using an annotation file of fQRS locations and then detected fetal QRS locations.

The above-reviewed methods use raw abdominal ECG signals to detect fetal QRS events. Unlike other studies, Lo et al. [4] used time-frequency representation to train a deep learning model. They used short-time-Fourier-transform (STFT) to prepare the features for training. Their deep learning model contains two-dimensional (2D) convolutional layers to extract spatial features from time-frequency images. They used a deep learning model to classify four classes (fetal QRS, maternal QRS, both, none) and yet the reported accuracy was around 90%. In the proposed method by Delgado et al. [20], Haar-Wavelet was applied to segments of AECG and then an interquartile range algorithm (IQR) was used to determine the difference between the four classes. Then, principal component analysis (PCA) was used to reduce signal dimensions. After this stage, they used three classifiers, namely k-nearest neighbour (kNN), support vector machine (SVM), and Bayesian network (BN). Zhang et al. [21] used template subtraction based on PCA to extract FECG from AECG signals. They preprocessed signals using a Butterworth bandpass filter in order to remove noises and artifacts. K-means clustering with the feature of max-min pairs was used to find fetal and maternal QRS. Finally, fQRS was extracted using template subtraction based on PCA.

Muduli et al. [11] used deep learning to extract fetal ECG. They developed an auto-encoder using an artificial neural network (ANN). Also, Zhong et al. [22] used deep learning in order to extract fetal ECG from a single-channel ECG signal directly. A residual convolutional encoder-decoder network was employed to extract features and estimate fetal ECG signals. In another work, Fotiadou et al. [12], [13] used a deep learning auto-encoder model on multi-channel data in order to denoise ECG signals. Fotiadou et al. [12], [13] used deep learning to extract FECG signals or denoise AECG signals, and they do not need the exact location of maternal or fetal QRS in ECG signals.

Xiong et al. [16] used a deep neural network (DNN) for denoising ECG signals. First, they used a wavelet transform (WT) method to filter out noise. Then, they used a DNN based on an auto-encoder to eliminate any remaining noise with an unknown distribution in the frequency domain that might be similar to ECG signal. They considered MIT-BIH dataset in this work where the proposed method has been shown to successfully remove noise from the signal. The average root mean square error (RMSE) obtained using this proposed method was reported to be less than 0.037. Zhao et al. [15] used a stacked long short-term memory network (Stacked-LSTM) to improve blind source separation performance. Their proposed model used an auto-encoder with a separation part which includes the stacked-LSTM with the attention mechanism of the squeeze-and-excitation (SE) module. It has been shown that this method has better performance in separating blind sources than classic methods. Mohebian et al. [14] used a generative adversarial network (GAN) to extract FECG from AECG. Their proposed model used an attention mechanism in order to attend to specific parts of ECG signal trying to suppress the maternal R wave instead of amplifying it. They used *sine* activation function for retaining more details after each convolution layer. They used the abdominal and direct FECG (A&D FECG) dataset to train their model and test it with Set-A of the 2013 Physionet/Computing in Cardiology Challenge dataset. It was shown that this model can extract FECG signal from the abdominal with a high F1-score (about 99.5%) and the results are comparable to other models.

III. Materials and Methods

In this paper, we propose an auto-encoder (AE) model equipped with a dual attention module to extract FECG segments from raw AECG. There are two major steps in the processing pipeline: *i*) pre-processing, and *ii*) FECG pattern extraction. For pre-processing, band-pass filtering, down-sampling, and sliding window are performed to prepare AECG data to feed to the network. Then, the proposed auto-encoder deep learning (AEDL) model is trained using the pre-processed data along with the existing labels; the expected output will be the extracted fetal ECG patterns.

A. Datasets

Three datasets, all of them belonging to PhysioNet[1], are used in this study. The first one is Non-invasive Fetal ECG (NIFECG1)–The PhysioNet Computing in Cardiology Challenge 2013 [23]. It includes 75 AECG recordings. The sampling rate of the signals is 1 KHz and each recording includes 4 channels and has 60 seconds of data. Similar to previous works with this dataset, we removed 7 (out of 75) recordings with inaccurate reference annotations from our experiments [24]. This dataset includes two sets of files including signals and annotation. The annotation file contains only the locations of events where fQRS occurred.

The second dataset is the Non-Invasive Fetal ECG Database (NIFECG2), containing 55 multichannel abdominal ECG (3 or 4 abdominal signals) recording from a single subject between 21 to 40 weeks of pregnancy [23]. In this dataset, the sampling rate is 1 KHz and recorded signals were filtered using a band-pass filter between 0.01 to 100 Hz. This dataset contains fetal QRS time samples as well. Fourteen sub-sets, including numbers 154, 192, 244, 274, 290,3 323, 368, 444, 597, 733, 746, 811, 826, 906, are selected because other research selected almost the same sub-sets and it facilitates the comparison [9], [25].

In the above-mentioned datasets, i.e., NIFECG1 and NIFECG2, no associated fetal ECG data are provided. Hence, NIFECG1 and NIFECG2 are normally used in the testing phase only, and to ensure the generalisability of the proposed model to different datasets with the ability to extract fetal ECG signals. In one of our experiments, we demonstrate the performance of FECG extraction against existing fQRS labels provided in these datasets.

The third dataset is called Abdominal and Direct Fetal ECG (A&D FECG) [23], [26]. This dataset contains FECG recordings related to 5 different women, between 38 and 41 weeks of gestation. There are five data channels in this dataset: four abdominal channels and one channel corresponding FECG obtained from the fetal scalp. It is noteworthy to mention that we used this FECG signal channel in this dataset as labels for training the proposed model. Each recording last for five minutes and the sampling frequency is 1 KHz. All signals were filtered using a band-pass filter between 1 and 150 Hz. Also, power-line interference (50Hz) and baseline drift are removed.

B. Pre-processing

In order to prepare the input data for the main processing, a Butterworth band-pass filter is applied to raw AECG signals. The cut-off frequencies of this filter are selected between 1 and 100 Hz [14]. This is to remove some high-frequency noises and produce a smooth signal. The filtered signal is then down-sampled to 200 Hz to reduce unnecessary computations and provide a universal data arrangement between both datasets. The filtered AECG signal is then segmented into small chunks before feeding to the network. The non-overlapping windows with size $W = 200$ samples (equivalent to 1 second) are selected empirically. The 1-second window is selected according to the maternal and fetal heartbeat rate to guarantee the inclusion of a sufficient number of ECG cycles inside a window to optimise the functionality of the proposed deep learning model when extracting FECG patterns. As an example, a segment of the filtered AECG (input to network) and the corresponding FECG signal is shown in Figure 1. Technically, the existing FECG signals are treated as labels to the proposed model during the training phase.

C. FECG pattern extraction

Auto-encoders are specific types of unsupervised feed-forward neural networks that follow the characteristics of input data to generate similar outputs. This type of neural network can deal with unlabelled data and learns how to efficiently compress and encode data. This ability is then complemented by learning how to reconstruct the output (from the reduced encoded representation) as close to the original input as possible. The main advantages of AEs are reducing dimensionality as well as compactness and speed. Their superiority over techniques such as multi-scale PCA [27] is that they are capable of modelling complex non-linear functions whereas PCA is essentially a linear transformation. This provides the flexibility to model fetal ECG components with the least influence from noises and maternal ECG mixture to the output signal. These features motivated us to base our proposed model on auto-encoders. AEs are commonly used to extract, construct, and reconstruct the input data in accordance with the application of interest, and are widely used in signal processing (denoising, extraction, etc.) [28]–[32].

Here, we use A&D FECG dataset to train the proposed auto-encoder. The reason for selecting this dataset for training is the availability of actual FECG signals along with their corresponding abdominal signals. The dataset contains two categories of signals: FECG, and AECG. The abdominal ECG signals generally contain three key components [33] which can be represented in form of an additive mixture, $\mathbf{s} = \mathbf{f} + \mathbf{m} + \omega$ where $\mathbf{s} = [s(1), s(2), ..., s(N)]$ is synthetic signal (AECG), $\mathbf{f} = [f(1), f(2), ..., f(N)]$ is FECG signal, $\mathbf{m} = [m(1), m(2), ..., m(N)]$ is MECG signals, $\omega = [\omega(1), \omega(2), ..., \omega(N)]$ is noise, and N is the number of

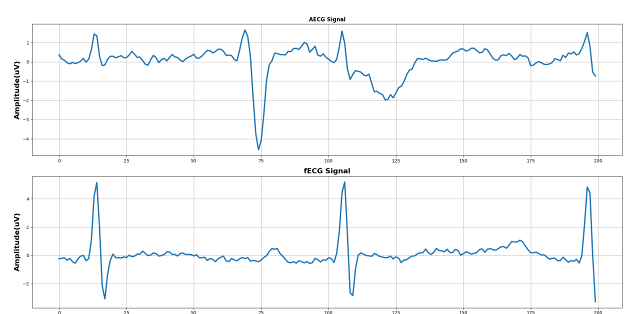


Fig. 1. Samples of raw AECG and FECG signals in A&D FECG dataset.

[1]https://physionet.org/about/database/

samples in the signal. The proposed model receives the AECG signals as input and tries to learn the related FECG patterns and extract them from the AECG signal. In other words, it is trained with AECG signals ($\mathbf{s}$) and extracts FECG signals ($\mathbf{f}$).

*1) **AEDL description**:* A generic AE consists of three main parts– encoder $e(.)$, code h, and decoder $d(.)$. The encoder compresses the input (s) and produces the code, $h = e(s)$. The generated code by the encoder is followed by a bidirectional-LSTM ($bi(.)$) Layer in order to create a feature map using past and future information, $h' = bi(h)$. The bidirectional-LSTM allows data flow in both directions, i.e. backwards or forward. In other words, the model can be trained using all available input information (past and future) to preserve the time order of the time-frames, which is essential for ECG pattern extraction [34]. The decoder attempts to reconstruct the desired output using the generated feature map h', $output = d(h')$ [35]. The loss function $L(x, output)$ is an important function in auto-encoder models to compare the input and the output of the model. During the training phase, the auto-encoder model attempts to minimise the loss function in order to find the output that best approximates the input signal.

One of the well-known types of AEs is fully connected (FC) networks. In this architecture, each neuron in the current layer is connected to all neurons in the previous and subsequent layers. This structure allows each neuron to extract a feature by observing all the possible connections in the previous layer [28], [36]. Unlike these models, convolutional-based auto-encoders (CNN-AEs) use some filters to extract local features from the input (using the kernel's convolution and parameter sharing). These networks usually have fewer parameters to be tuned compared to other CNNs, especially fully-connected neural networks. As a result, CNN-AE networks are computationally less expensive and require less memory storage [28]. CNN-AEs use local features each containing partial information about the input data. This provides sufficient flexibility to the model to choose the part of the data that contains the information of interest. In contrast, global features contain information on all input data in one place which is challenging to be divided. Due to these characteristics, we have chosen a CNN-AEs model for extracting FECG signals in this paper.

As reviewed earlier in this paper, attention modules have been recently utilised in the development of neural networks to provide higher performance. A proper attention module can help the auto-encoders to attend to specific features and construct the desired output. To the best of our knowledge, this is the first study where the benefits of using a dual attention module are explored within an auto-encoder model for FECG extraction. Next, we describe the definition of the attention module followed by the associated details.

*2) **Attention**:* Attention plays an important role in human cognition. In psychology, attention is the cognitive process of selectively concentrating on one or a few things while ignoring others. A neural network is considered to be an effort to mimic human brain actions in a simplified manner. The attention mechanism is also an attempt to implement the same action of selectively concentrating on a few relevant things while ignoring others in deep neural networks [37]. Attention mechanisms are inseparable parts of deep learning models which work with sequences. These mechanisms make it possible to model dependencies regardless of their distance in the input and output sequences [38], [39]. In this paper, a dual attention module is incorporated into an AE network to increase the performance of FECG pattern extraction. In the proposed model, attention modules learn how to find FECG cycles buried in the AECG signals and try to attend to FECG cycles and ignore the rest part of the AECG signals. The auto-encoder model can construct FECG signals using the selected FECG patterns by attention modules. These attention modules are described next.

- **Squeeze-and-Excitation (SE) module:** Squeeze-and-Excitation can be used as part of a CNN architecture to improve the ECG channels' interdependencies with negligible computational cost. Considering the channel-wise information plays a critical role in enhancing the discrimination of the latent embedding features, we draw inspiration from the SE module, which proves to benefit many vision tasks via capturing interdependencies among channels to selectively highlight informative features and suppress useless ones [40]. This appendment can be done with no significant structural distortion to CNN. This module learns how to attend to useful global information (FECG cycles) and to ignore useless information associated with each ECG channel (the residual parts of the AECG signals). Previously, Zhao et al. [15] used the SE module in their work to improve blind source separation performance. In another work, Li et al. [41] used SE attention in their work for heartbeat recognition. Both studies have used SE as part of a CNN architecture. The detailed diagram of the proposed SE Attention module used in this study is shown in Figure 2. This module consists of three main parts: Squeeze, Excitation, and Scale. The SE attention procedure starts with a global average pooling to squeeze the feature map. The feature map is extracted from the input AECG signal (through the previous convolution layer) and has the dimension $1 \times 1 \times C$. Parameter C ($C = 32$ selected empirically) denotes the number of the previous convolutional layer filters which depends on the size of the filter used in the previous convolution layer. Afterwards, two FC layers are implemented. The first FC layer is used to compress the feature map with a smaller C. This is done by multiplying an arbitrary $0 < \epsilon \leq 1$ to C, i.e, $C \times \epsilon$. The output of the second FC layer, followed by a Sigmoid activation function, would have the same size as C. This output can be used as new weights applied back to the original feature map. This process will scale each channel according to its importance by multiplying the mentioned feature map by the input data.
- **Channel-Wise (CW) module:** Channel-Wise Attention (sometimes called Self-Attention) connects different positions of a sequence together in order to find a meaningful representation of that sequence. This framework integrates the channel-wise attention mechanism into a CNN to explore spatial information, which can take the importance of different channels by channel-wise attention and the spatial information of multichannel

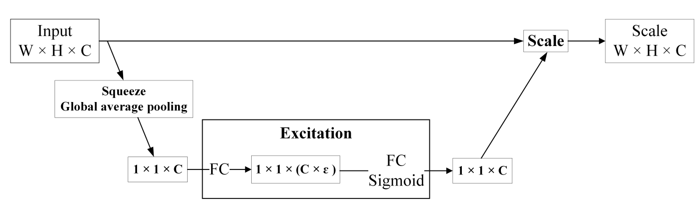


Fig. 2. SE attention module.

signals by a CNN into consideration [42], [43]. In general, this module is like another attention module in structure (SE attention mentioned above) but there is a small difference. In particular, CW attention can perform promisingly here because it can modify the information within the extracted feature map using the previous convolution layer in order to discover the weights of different AECG channels. This leads to the accurate extraction of important information from each AECG channel. So it can help the model to mitigate the information loss of input AECG. Previously, Srivastava et al. [44] used channel-wise attention in their model to detect multiple cardiac abnormalities using the ECG signal. Generally, this module can squeeze the global spatial information and generate channel-wise statistics. The advantage of this module is that it can be easily integrated and trained with CNN models. To the best of our knowledge, this is the first study in which CW is used for FECG extraction. CW consists of two fully connected layers where the first FC layer is followed by the hyperbolic tangent activation function ($\tanh$) and the second one is followed by Softmax activation function.

*3) **Dual Attention AEDL**:* In the proposed AEDL model, CNN extracts key features within the encoder parts using five convolutional layers. Then, the decoder tries to reconstruct the FECG patterns using extracted features through the convolutional layers. The key characteristic in the proposed model is the utilisation of two attention modules within both encoder and decoder parts, where a bidirectional-LSTM is used in between. The SE module is employed after the first convolution layer (within the encoder part) and the CW module is employed just before the last convolution layer (within the decoder part). The entire architecture of our proposed model is shown in Figure 3. All the convolutional layers in the proposed model include subsampling operations. During the training phase, the AECG mixture signals ($\mathbf{s}$) and FECG signals ($\mathbf{f}$) are fed to the encoder block as the input and label data, respectively. This block extracts features from AECG signals using a convolution layer which is mathematically represented via:

$$fm = g(\mathbf{W} * \mathbf{s}_f + b) \tag{1}$$

where fm denotes the generated feature map using a convolution layer, g is convolution's activation function which is ELU function in all convolution layers, $\mathbf{W}$ is the weight matrix connecting the input layer and hidden layer, b is bias, and $*$ denotes the convolution operation. These feature maps maintain the key components of the AECG signal. The SE attention module then amplifies useful global extracted features (the position of FECG cycles in the entire AECG signals) from the convolution layer through:

$$se_o = Sigmoid(\rho(\mathbf{W_2} \times \rho(\mathbf{W_1} \times GAP(fm) + b_1) + b_2)) \tag{2}$$

and then the output of the SE module is obtained via:

$$se_{attention} = fm \times se_o \tag{3}$$

where GAP denotes Global Average Pooling, ρ is the SE attention activation function (which is ReLU here), and se_o is the SE attention module output. All convolution layers are followed by ELU activation function in (4), where α is a value between 0 and 1, e is exponential sign, and z is summation of the weights (w) and the inputs (a) plus bias (b) as ($\Sigma_i a_i w_i + b$). As seen from Figure 3, a bidirectional-LSTM layer is used between the encoder and the decoder blocks. Bidirectional-LSTM layer is obtained by putting two independent LSTMs together. This structure allows the networks to have both backward and forward information about the sequence at every time step. A bidirectional-LSTM with no backward state is simplified to a unidirectional-LSTM. Similarly, a bidirectional-LSTM without the forward state will act as a unidirectional-LSTM with a reversed time axis [34]. During the encoding process, and because of subsampling, the input AECG signal information might be lost. Hence, we adopt a CW attention module (as detailed above) in the decoding block to mitigate this drawback. In the following, the decoder block attempts to reconstruct missing data. In other words, the decoder tries to reconstruct the input signal using the extracted temporal features in the bidirectional-LSTM layer, i.e., fm_b. The main part in the decoder block is the Channel-Wise attention module inserted before the final convolution layer (Figure 3). This module determines the most effective features which can help to construct FECG signals. It generates a probabilistic distribution of channels and then multiplies it by input to determine the most effective channels. The CW attention operation is summarised in the following equations:

$$R(z) = \begin{cases} z, & z > 0 \\ \alpha.(e^z - 1), & z \leq 0 \end{cases} \tag{4}$$

$$cw_o = Softmax(\mathbf{W_2} \times tanh(\mathbf{W_1} \times fm' + b_1) + b_2) \tag{5}$$

$$cw_{attention} = fm' \times cw_o \tag{6}$$

where $\mathbf{W}$ is the weight matrix connecting neurons of different layers, b is bias, fm' is the generated feature map from the previous convolution layer, and cw_o is the probabilistic distribution of the input feature map channels. the probabilistic distribution generated from CW attention module cw_o is multiplied to the input feature map fm' as shown in (6). Finally, the last convolution layer constructs final output (i.e., FECG signals).

*4) **Model parameters**:* As shown in Figure 3, both encoder and decoder blocks consist of 5 convolution layers. Each layer is followed by an ELU activation function. The filter parameters in the convolution layers in the encoder part are {32, 64, 128, 256, 512}. The number of neurons of fully-connected layers in the SE attention module is {32, 32} followed by ELU activation function. Then, the output of the Sigmoid activation function is multiplied by the input and fed to the dropout layer with a probability of 0.5. After the encoder part, a bidirectional-LSTM layer with 512 cells is used. The filter parameters associated with the convolution layers in the decoder part are {256, 128, 64, 32, 1} values. The number of neurons of fully-connected layers in the Channel-Wise attention module is {32, 32}. The first fully-connected layer is followed by the $\tanh$ activation function and the second one is followed by Softmax activation function. Then, the output of the Softmax activation function is multiplied by the input and fed to the last convolution layer in order to construct the final result. Adam and mean squared error (MSE) are used for optimizer and loss function, respectively. The model parameters were selected empirically as done in most relevant studies. To ensure the selection of the best parameters, we ran the model many times with various parameter values like learning rate, the number of neurons in each layer,

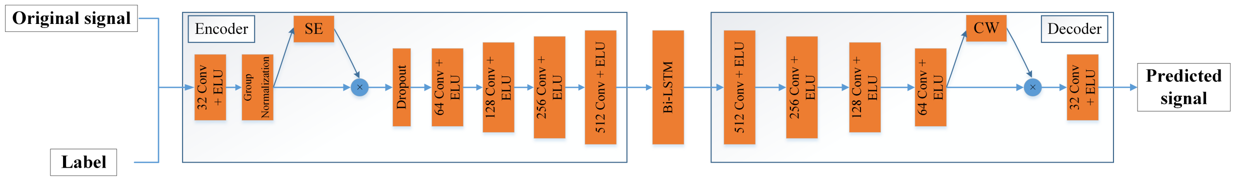


Fig. 3. The architecture of proposed Auto-Encoder deep learning model.

activation functions, the place of attention modules, the batch size and epochs. The trend of model learning and accuracy was analysed after each run and the best parameter settings were recorded.

IV. Experimental results

We have conducted several experiments to evaluate the performance of the proposed model. An ablation study is considered to assess the impact of different layers and parameters on the performance of the proposed model. To do this, we compare our proposed model with two different variants. The first variant (called MODEL-I) is an auto-encoder model with a bi-directional LSTM layer only. The attention modules are removed in this model. MODEL-I contains 5 convolutional layers for the encoding part, followed by a bi-directional LSTM to extract temporal features. Then, 5 convolutional layers are used in the decoding part to generate the FECG signals.

The second variant (called MODEL-II) is a simple auto-encoder model - using 5 convolutional layers for the encoding part and 5 convolutional layers for the decoding part. Both the bi-directional LSTM layer and attention modules are removed in this model. These two models are structured as an ablation study to show the impact of different layers in the proposed architecture. We intentionally kept the parameters unchanged in different models to be able to observe the effects of the existence of various layers/modules under similar conditions.

To run the experiments, we first apply the pre-processing steps to all the available AECG signals. Then, A&D FECG dataset is divided into train, test, and validation sub-sets, randomly. The optimizer and loss function considered in all experiments are Adam and Mean Squared Error (MSE), respectively. Kernel size and strides for all convolution layers in the proposed model (shown in Figure 3) are set to 3 and 1, respectively. The learning rate is set to 10^{-4}.

Figure 4 illustrates sample results for a single subject from the test sub-set in A&D FECG dataset. As shown in Figure 4, both variants (MODEL-I, MODEL-II), as well as the proposed model, can extract FECG signals. However, the proposed model outperforms MODEL-I and MODEL-II. This is very well observed from Figure 4 - right column. Among these models, MODEL-II (with no attention modules and no bi-directional LSTM layer) performs poorly and MODEL-I performs slightly better. This experiment clearly shows the effects of adding attention modules and a bi-directional LSTM layer to the deep model. In fact, the classic auto-encoder model (MODEL-II) cannot reconstruct FECG signals satisfactorily as shown in Figure 4. As expected, the bidirectional-LSTM layer improved the performance of the auto-encoder model to extract and learn FECG temporal features and then construct FECG signals.

Figure 5 depicts the extraction results using three models in two different subjects (Subjects 1 and 33) from NIFECG1 dataset. In this figure, blue signals are AECG, and red ones are extracted FECG signals. As mentioned in Section III-A, NIFECG1 does not have the FECG ground truth. hence, the blue stars in this figure, overlayed on red signals, show the corresponding FECG R-peaks. As shown in this figure, MODEL-II can not find fetal QRS nor extract FECG signal but two other models, i.e., MODEL-I and the proposed model, can construct FECG signal clearly. The constructed FECG signal using MODEL-I matches fetal QRS, much better than in MODEL-II. However, such cross-matching is far better in the proposed model with a very high amplitude. Subject 01 in Figure 5 is a good sample for illustrated the proposed method performance. There are 3 fetal QRS in this sample one of which is near to maternal QRS (the first one) and one of them overlaps with maternal QRS (the last one). As demonstrated in this figure, our model could extract fetal QRS. Another point observed in this subject is about the second fetal QRS. As seen from Figure 5, despite the low amplitude of Fetal QRS in AECG signals and the dominance of maternal ECG signals, the amplitude of the extracted fetal ECG using the proposed method is considerably high. The main reason for this observation is that the FECG signals used as labels to train the model (Figure 1) have high amplitudes. Hence, the extracted fetal ECG for Subject 01 in Figure 5, is also obtained with relatively higher amplitude than what was observed in the AECG mixture.

To further explore the behaviour of these variants, loss values in each epoch of the training phase for all models are calculated via:

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{7}$$

where Y and $\hat{Y}$ are original FECG and predicted FECG, respectively. And N is the number of data samples. Figure 6 shows the decreasing trend of the loss function for each ablation setting. All models are trained in 100 epochs. Figure 6 shows that the best loss decay trend belongs to the proposed model. Also, it is seen that MODEL-II behaves far worse than the other two models verifying the key roles of using both attention modules and the bi-directional LSTM layer. We also used two other datasets (NIFECG1 and NIFECG2) to estimate the performance of the proposed model in order to be able to compare the results with related models. The proposed model was trained using A&D FECG dataset and then estimated using NIFECG1 and NIFECG2 separately. F1-measure and accuracy are used as performance metrics. In the following, the impact of using attention modules and the bi-directional LSTM layer are investigated numerically. As shown in Table I and Figures 7 and 8, a sole auto-encoder cannot extract FECG from AECG signal so that the fetal QRS are detectable. The accuracies of this model for NIFECG1 and NIFECG2 are 83.29% and 77.22%, respectively. Adding a bidirectional LSTM layer has considerably increased the accuracy. This layer has improved the accuracy by about 8% and 6% in NIFECG1 and NIFECG2, respectively. In the proposed model

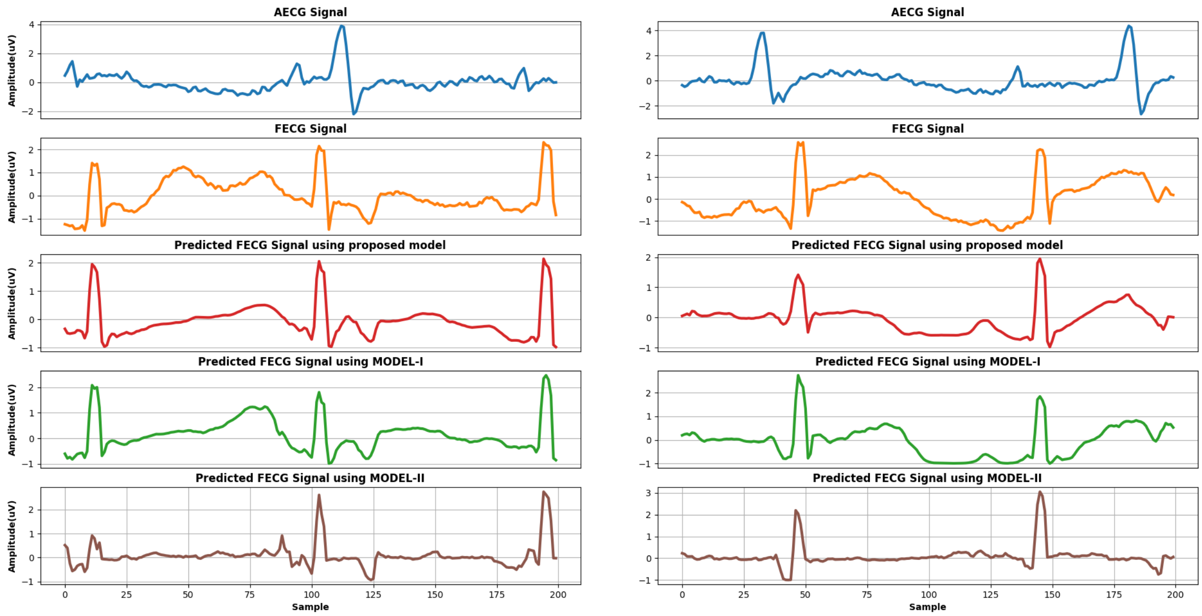


Fig. 4. Two examples of FECG generation using three proposed models for subject 1 from the test sub-set in A&D FECG dataset. In each subplot, the top one is the AECG signals, the second signals are original FECG, the third signals are generated FECG signals using the proposed model, the fourth signals are generated FECG signals using MODEL-I, and the last signals are generated FECG signals using MODEL-II. All signals are on 200 samples of test set from A&D FECG dataset.

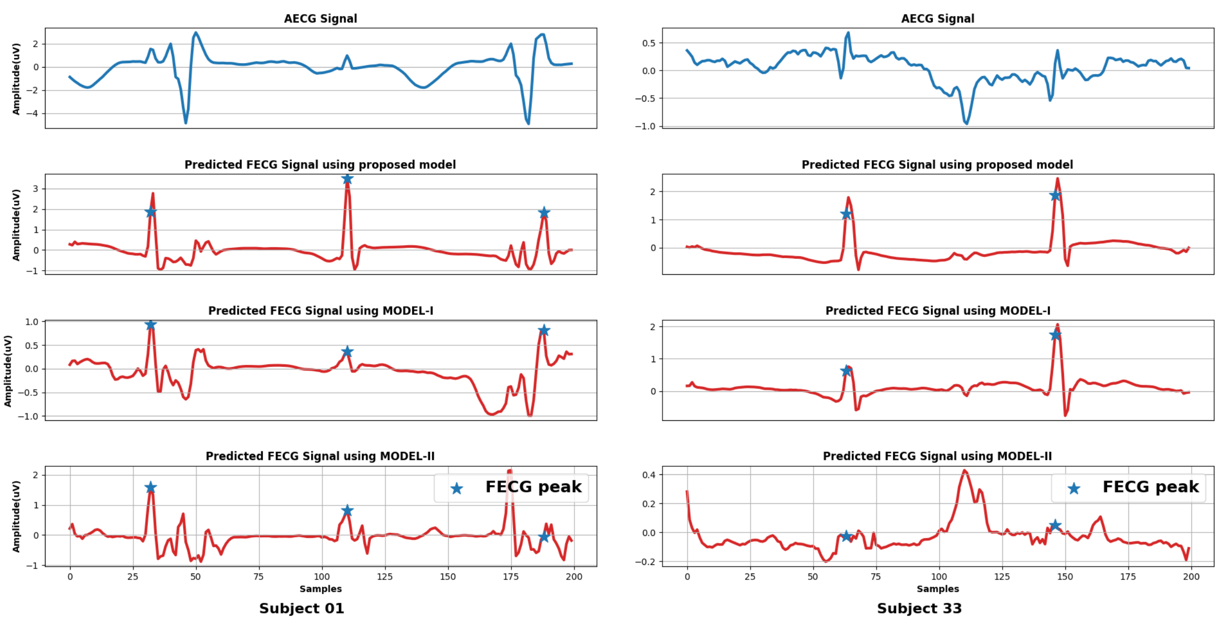


Fig. 5. Two examples of FECG generation using three proposed models in the test dataset (NIFECG1 dataset). In each subplot, the top one is the AECG signals, The second signals are original FECG, the third signals are generated FECG signals using the proposed model, the fourth signals are generated FECG signals using MODEL-I, and the last signals are generated FECG signals using MODEL-II. The star shapes in all plots are the real fetal QRS. All signals are on 200 samples of the test set in NIFECG1 dataset.

- AEDL - two attention modules are used in order to attend to the fetal ECG patterns in ECG signals. These two attention modules improved the accuracy by about 5% in both datasets and the accuracies are now 96.26% and 88.54% for NIFECG1 and NIFECG2, respectively. According to the obtained results, the proposed auto-encoder model provides very clean morphological features within the extracted FECG signal. This paves the path for using the proposed method in practical medical diagnosis. For further quantitative evaluations, the F1-measure, the sensitivity (SE), and the positive predictive value (PPV) are calculated via:

$$SE = \frac{TP}{TP + FN} \tag{8}$$

$$PPV = \frac{TP}{TP + FP} \tag{9}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FN + FP} = 2 \times \frac{SE \times PPV}{SE + PPV} \tag{10}$$

where TP, FN, and FP are true positive, false negative, and false positive, respectively. The F1-measures for all models are indicated in Figure 7 and Table I.

TABLE I
THE COMPARISON OF FQRS DETECTION ACCURACY, F1-MEASURE, SENSITIVITY, AND PPV METRICS AS A RESULT OF APPLYING THE THREE METHODS TO 69 OUT OF THE TOTAL NUMBER OF SUBJECTS IN THE TEST DATASET NIFECG1 AND 14 OUT OF THE TOTAL NUMBER OF SIGNALS IN TEST DATASET NIFECG2.

Dataset	Method	Acc(%) $\pm$ std	F1-score(%) $\pm$ std	Sensitivity(%) $\pm$ std	PPV(%) $\pm$ std
NIFECG1	MODEL-II	83.30 $\pm$ 7.25%	89.92 $\pm$ 4.81%	87.36 $\pm$ 5.86%	92.66 $\pm$ 3.65%
	MODEL-I	91.61 $\pm$ 4.88%	95.33 $\pm$ 2.91%	93.95 $\pm$ 3.69%	93.42 $\pm$ 3.10%
	Proposed model	**96.26** $\pm$ **2.86**%	**98.02** $\pm$ **1.58**%	97.38 $\pm$ 2.05%	98.68 $\pm$ 1.14%
NIFECG2	MODEL-II	77.22 $\pm$ 15.51%	84.35 $\pm$ 13.24%	81.43 $\pm$ 14.52%	87.61 $\pm$ 11.60%
	MODEL-I	83.22 $\pm$ 10.01%	89.57 $\pm$ 7.17%	87.10 $\pm$ 8.45%	90.05 $\pm$ 6.65%
	Proposed model	**88.54** $\pm$ **11.17**%	**92.87** $\pm$ **8.60**%	91.21 $\pm$ 9.84%	94.65 $\pm$ 7.12%

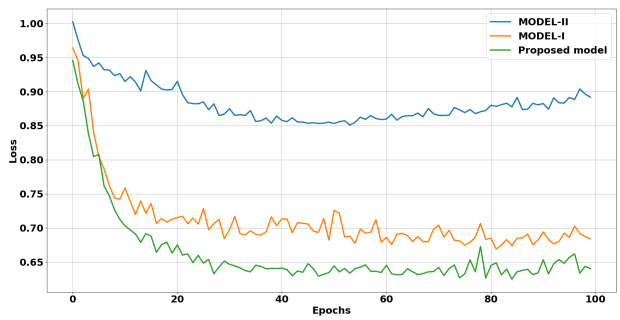


Fig. 6. RMS loss value for each epoch in three different models.

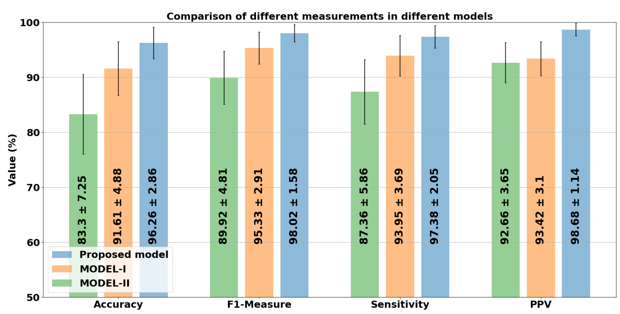


Fig. 7. The average and standard deviation on the accuracy, F1-measure, Sensitivity, and PPV metrics as a result of applying the three methods to 69 out of the total number of subjects in the test dataset NIFECG1.

For a detailed assessment, Table II shows a comparison between proposed models and the state-of-the-art for fetal QRS detection. In this table, several methods working based on both deep learning and traditional classification are considered. In this work, attempts have been made to use the same datasets for training and testing models in order to show our models' robustness. As shown in Table II, our proposed method has a superior performance with dataset NIFECG1. However, the results with dataset NIFECG2 are slightly lower. This can be due to (as mentioned in section III-A) NIFECG2 being recorded using subjects between 21 to 40 weeks of pregnancy. The FECG signal in the first few weeks would be very weak and fetal QRS is combined with noise. In this dataset, as we get closer to the final pregnancy weeks, the recorded signals become stronger and hence the fetal signal would be more recognizable. Nevertheless, the proposed model can construct FECG signals very well. Among these, only Varanini et al. [45] performs slightly better than our method with NIFECG1 dataset.

From Table II we can see that Behar et al. [9] used template subtraction (TS), recurrent neural network (RNN), PCA, LMS, and RLS in order to evaluate their model to detect fetal QRS based on P&T method. They used 30 seconds of ECG signals in NI-FECG2 dataset and used an overlapping window sized

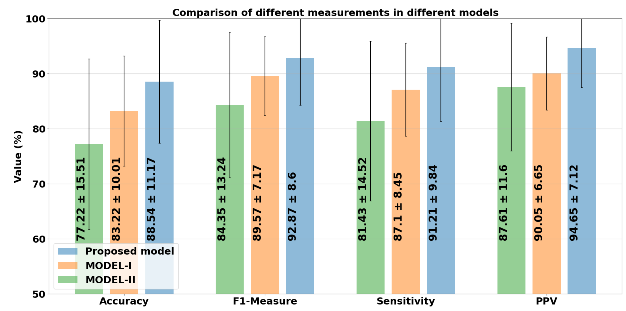


Fig. 8. The average and standard deviation on the accuracy, F1-measure, Sensitivity, and PPV metrics as a result of applying the three methods to 14 out of the total number of signals in the test dataset NIFECG2.

50 ms to train their models and they evaluated their model on the external dataset. Nevertheless, the proposed model in our study obtained better results. Our results are also superior to the Encoder-Decoder method, proposed in [22], where the same dataset was used to train the model. Zhang et al. [46] used SVD-SW and they achieved better accuracy than the proposed model on A&D FECG dataset. But they do not evaluate their model using an external dataset. Behar et al [47] used the combination of ICA and template subtraction trained on NIFECG1. The proposed model also performed better in this case. Varanini et al. [45] used ICA and a specialized post-processing method in order to extract fetal ECG signals and detect fetal QRS. They trained their model on NIFECG1 dataset and like the work in [46], the performance is higher than the proposed model but again they did not evaluate their model on an external dataset. Mohebbian et al. [14] proposed a new deep learning model in order to extract fetal ECG signals. They proposed a generative adversarial network (GAN). Their model performed well but our proposed model outperformed Mohebbian et al. [14]. Karvounis et al. [48] used a three-stage approach called multi-scale PCA (MSPCA) followed by Smoothed Nonlinear Energy Operator (SNEO). They used these two methods in order to detect maternal QRS and extract fetal QRS that do not overlap with maternal QRS. The multi-scale PCA is used for de-noising and SNEO extract maternal and fetal QRS. They did not report the F1-measure but the accuracy is 95% which underperforms the proposed method.

Through the results provided in Table III, we investigate the impact of selecting different window sizes on the overall performance of the proposed model. Since window size (W) can have a significant impact on the model's complexity, the computation times (training phase) are also reported. According to Table III, the accuracy and F1-measure drop for small Ws. Most notable decline occurs at $W = 100$ (equivalent to 0.5 ms). This is because very small windows do not allow the neural network to capture attention between full

TABLE II
THE COMPARISON OF FQRS DETECTION WITH STATE-OF-THE-ART.

Method	F1-score(%)	Dataset	QRS detection method	Number of tested signals
TS and ES-RNN [9]	97.2	NIFECG2	P&T	14
	90.2	Train on NIFECG2 and test on private dataset		
TS and PCA [9]	95.4	NIFECG2		
	89.3	Train on NIFECG2 and test on private dataset		
LMS [9]	95.4	NIFECG2		
	87.9	Train on NIFECG2 and test on private dataset		
RLS [9]	95.9	NIFECG2		
	88.2	Train on NIFECG2 and test on private dataset		
OBACKC [25]	95.6	NIFECG2	Thresholding	
SVD-SW [46]	99.4	A&D FECG	P&T	2
Encoder-Decoder [22]	94.1	A&D FECG		5
Behar et al. [47]	95.9	NIFECG1		69
Varanini et al. [45]	99.0	NIFECG1	Customized	69
Mohebbian et al. [14]	94.7	Train on A&D FECG and test on NIFECG2	P&T	14
	97.9	Train on A&D FECG and test on NIFECG1		69
MODEL-I **MODEL-II** **Proposed model**	86.2 90.5 93.5	Train on A&D FECG and test on NIFECG2	Hamilton algorithm	14
MODEL-I **MODEL-II** **Proposed model**	90.7 95.5 98.1	Train on A&D FECG and test on NIFECG1	P&T	69

cycles of FECG and MECG. As the window size increases, the performance improves which consequently increases the computation time. According to the literature, $W = 200$ (equivalent to 1s) would ensure the presence of at least one cycle of ECG pattern (QRST) for the network to be trained. This is verified by the results in Table III where the accuracy is not significantly improved for $W > 200$. For example, the accuracy difference between $W = 500$ and $W = 200$ is about 0.8%. It is therefore rational to select $W = 200$ (equivalent to 1s) throughout the experiments to meet a balance in accuracy and complexity. This also sets a common ground for a fair comparison between different related methods in Table II.

V. CONCLUSION

In this study, fetal ECG extraction from abdominal ECG using a novel deep learning model was addressed. The proposed model is composed of an auto-encoder with a dual attention module and a bidirectional LSTM layer. One of the major advantages of the proposed model is no need for prior information about maternal ECG. It also takes the advantage of a bi-directional LSTM layer which allows for preserving prior and posterior information at any point of the input signal. This leads to finding (and extracting) the most relevant pattern to FECG signal. Furthermore, the proposed dual attention mechanism provides a mask to certain parts of the signal and avoids processing parts that can increase errors. It allows focusing on one part of the input (FECG in our case) while giving less attention to others, thus guiding the process of reasoning. The disadvantages are that the model's performance drops with signals with a low signal-to-noise rate or when the fetal QRS is too weak. Furthermore, the network needs to be optimised/compressed in order to be used in real-time and/or on edge devices. Three large datasets were considered to evaluate the performance of the proposed model. The results of our extensive experiments with these datasets confirm the effectiveness and generalisability of the proposed methods. The obtained results were superior to those recorded by the state-of-the-art.

REFERENCES

[1] D. Gurve and S. Krishnan, "Separation of Fetal-ECG From Single-Channel Abdominal ECG Using Activation Scaled Non-Negative Matrix Factorization," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 669–680, 2019.

[2] S. M. Gilboa, O. J. Devine, J. E. Kucik, M. E. Oster, T. Riehle-Colarusso, W. N. Nembhard, P. Xu, A. Correa, K. Jenkins, and A. J. Marelli, "Congenital heart defects in the united states: estimating the magnitude of the affected population in 2010," *Circulation*, vol. 134, no. 2, pp. 101–109, 2016.

[3] R. Sameni and G. D. Clifford, "A review of fetal ecg signal processing; issues and promising directions," *The open pacing, electrophysiology & therapy journal*, vol. 3, p. 4, 2010.

[4] F.-W. Lo and P.-Y. Tsai, "Deep learning for detection of fetal ecg from multi-channel abdominal leads," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2018, pp. 1397–1401.

[5] R. Kahankova, R. Martinek, R. Jaros, K. Behbehani, A. Matonia, M. Jezewski, and J. A. Behar, "A review of signal processing techniques for non-invasive fetal electrocardiography," *IEEE reviews in biomedical engineering*, vol. 13, pp. 51–73, 2019.

[6] H. Wan, Q. Liu, and J. Chai, "A method for extracting fecg based on ica algorithm," in *2008 9th International Conference on Signal Processing*. IEEE, 2008, pp. 2761–2764.

[7] S. Nikam and S. Deosarkar, "Fast ica based technique for non-invasive fetal ecg extraction," in *2016 Conference on Advances in Signal Processing (CASP)*. IEEE, 2016, pp. 60–65.

[8] D. Gurve and S. Krishnan, "Separation of fetal-ecg from single-channel abdominal ecg using activation scaled non-negative matrix factorization," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 669–680, 2020.

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

[10] W. Zhong, L. Liao, X. Guo, and G. Wang, "A deep learning approach for fetal qrs complex detection," *Physiological measurement*, vol. 39, no. 4, p. 045004, 2018.

[11] P. R. Muduli, R. R. Gunukula, and A. Mukherjee, "A deep learning approach to fetal-ecg signal reconstruction," in *2016 Twenty Second National Conference on Communication (NCC)*. IEEE, 2016, pp. 1–6.

[12] E. Fotiadou, T. Konopczyński, J. Hesser, and R. Vullings, "Deep convolutional encoder-decoder framework for fetal ecg signal denoising," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. Page–1.

[13] E. Fotiadou and R. Vullings, "Multi-channel fetal ecg denoising with deep convolutional neural networks," *Frontiers in Pediatrics*, vol. 8, 2020.

[14] M. R. Mohebian, S. shahim Vedaei, K. A. Wahid, A. Dinh, H. R. Marateb, and K. Tavakolian, "Fetal ecg extraction from maternal ecg using attention-based cyclegan," *IEEE Journal of Biomedical and Health Informatics*, 2021.

TABLE III

THE COMPARISON OF FQRS DETECTION ACCURACY, F1-MEASURE, SENSITIVITY, AND PPV METRICS AS A RESULT OF APPLYING THE PROPOSED METHOD (DUAL ATTENTION AEDL) WITH DIFFERENT WINDOW SIZES TO 69 SUBJECTS IN THE TEST DATASET NIFECG1. TIME COLUMNS (AVERAGE TIME PER EPOCH AND TOTAL TIME) ARE IN SECONDS AND DEMONSTRATE THE TIME THAT ONE WINDOW TAKES FROM THE MODEL.

Dataset	Window Size (W)	Acc(%) $\pm$ std	F1-score(%) $\pm$ std	Sensitivity(%) $\pm$ std	PPV(%) $\pm$ std	Time per epoch (s)	Total time (s)
NIFECG1	500	97.83 $\pm$ 2.41%	98.86 $\pm$ 1.30%	98.49 $\pm$ 1.69%	99.24 $\pm$ 0.94%	0.054	1.633
	400	97.52 $\pm$ 2.80%	98.69 $\pm$ 1.52%	98.26 $\pm$ 1.97%	99.13 $\pm$ 1.10%	0.044	1.331
	300	96.98 $\pm$ 3.01%	98.40 $\pm$ 1.64%	97.94 $\pm$ 2.13%	98.87 $\pm$ 1.18%	0.033	0.994
	200	96.26 $\pm$ 2.86%	98.02 $\pm$ 1.58%	97.38 $\pm$ 2.05%	98.68 $\pm$ 1.14%	0.022	0.669
	100	93.53 $\pm$ 4.97%	96.43 $\pm$ 2.92%	95.41 $\pm$ 3.70%	97.50 $\pm$ 2.11%	0.011	0.333

[15] M. Zhao, X. Yao, J. Wang, Y. Yan, X. Gao, and Y. Fan, "Single-channel blind source separation of spatial aliasing signal based on stacked-lstm," *Sensors*, vol. 21, no. 14, p. 4844, 2021.

[16] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "Ecg signal enhancement based on improved denoising auto-encoder," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 194–202, 2016.

[17] J. S. Lee, M. Seo, S. W. Kim, and M. Choi, "Fetal qrs detection based on convolutional neural networks in noninvasive fetal electrocardiogram," in *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*. IEEE, 2018, pp. 75–78.

[18] K. Vo, T. Le, A. M. Rahmani, N. Dutt, and H. Cao, "An efficient and robust deep learning method with 1-d octave convolution to extract fetal electrocardiogram," *Sensors*, vol. 20, no. 13, p. 3757, 2020.

[19] A. Huque, K. Ahmed, M. Mukit, and R. Mostafa, "Hmm-based supervised machine learning framework for the detection of fecg rr peak locations," *IRBM*, vol. 40, no. 3, pp. 157–166, 2019.

[20] J. A. Delgado, M. Altuve, and M. N. Homsi, "Haar wavelet transform and principal component analysis for fetal qrs classification from abdominal maternal ecg recordings," in *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*. IEEE, 2015, pp. 1–6.

[21] Y. Zhang and S. Yu, "Single-lead noninvasive fetal ecg extraction by means of combining clustering and principal components analysis," *Medical & biological engineering & computing*, vol. 58, no. 2, pp. 419–432, 2020.

[22] W. Zhong, L. Liao, X. Guo, and G. Wang, "Fetal electrocardiography extraction with residual convolutional encoder–decoder networks," *Australasian physical & engineering sciences in medicine*, vol. 42, no. 4, pp. 1081–1089, 2019.

[23] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[24] J. Behar, J. Oster, and G. D. Clifford, *Non-invasive FECG extraction from a set of abdominal sensors*. IEEE, 2013.

[25] 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," *Biomedical Signal Processing and Control*, vol. 57, p. 101673, 2020.

[26] 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," *Biomedizinische Technik/Biomedical Engineering*, vol. 57, no. 5, pp. 383–394, 2012.

[27] A. A. Akinduko and A. N. Gorban, "Multiscale principal component analysis," *Journal of Physics: Conference Series*, vol. 490, p. 012081, mar 2014.

[28] W. Li, H. Xu, Z. Ma, R. Zhu, D. Hu, Z. Zhu, J. Gu, C. Shan, J. Zhu, and X.-P. Wu, "Separating the eor signal with a convolutional denoising autoencoder: a deep-learning-based method," *Monthly Notices of the Royal Astronomical Society*, vol. 485, no. 2, pp. 2628–2637, 2019.

[29] T. Wen and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals," *IEEE Access*, vol. 6, pp. 25 399–25 410, 2018.

[30] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ecg signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60 806–60 813, 2019.

[31] S. Nagar, A. Kumar, and M. Swamy, "Orthogonal features-based eeg signal denoising using fractionally compressed autoencoder," *Signal Processing*, vol. 188, p. 108225, 2021.

[32] S. Nagar and A. Kumar, "Orthogonal features based eeg signals denoising using fractional and compressed one-dimensional cnn autoencoder," *arXiv preprint arXiv:2104.08120*, 2021.

[33] J. Behar, F. Andreotti, S. Zaunseder, J. Oster, and G. D. Clifford, "A practical guide to non-invasive foetal electrocardiogram extraction and analysis," *Physiological measurement*, vol. 37, no. 5, p. R1, 2016.

[34] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[35] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[36] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

[37] J. Essington, "How we learn: why brains learn better than any machine . . . for now," *Educational Review*, vol. 0, no. 0, pp. 1–1, 2021.

[38] C. Raffel, M.-T. Luong, P. J. Liu, R. J. Weiss, and D. Eck, "Online and linear-time attention by enforcing monotonic alignments," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2837–2846.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[40] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[41] X. Li, F. Zhang, Z. Sun, D. Li, X. Kong, and Y. Zhang, "Automatic heartbeat classification using s-shaped reconstruction and a squeeze-and-excitation residual network," *Computers in biology and medicine*, vol. 140, p. 105108, 2022.

[42] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5659–5667.

[43] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "Eeg-based emotion recognition via channel-wise attention and self attention," *IEEE Transactions on Affective Computing*, 2020.

[44] A. Srivastava, A. Hari, S. Pratiher, S. Alam, N. Ghosh, N. Banerjee, and A. Patra, "Channel self-attention deep learning framework for multi-cardiac abnormality diagnosis from varied-lead ecg signals," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.

[45] M. Varanini, G. Tartarisco, L. Billeci, A. Macerata, G. Pioggia, and R. Balocchi, "An efficient unsupervised fetal qrs complex detection from abdominal maternal ecg," *Physiological measurement*, vol. 35, no. 8, p. 1607, 2014.

[46] N. Zhang, J. Zhang, H. Li, O. O. Mumini, O. W. Samuel, K. Ivanov, and L. Wang, "A novel technique for fetal ecg extraction using single-channel abdominal recording," *Sensors*, vol. 17, no. 3, p. 457, 2017.

[47] J. Behar, J. Oster, and G. D. Clifford, "Combining and benchmarking methods of foetal ecg extraction without maternal or scalp electrode data," *Physiological measurement*, vol. 35, no. 8, p. 1569, 2014.

[48] E. C. Karvounis and D. I. Fotiadis, "Maternal and fetal heart rate extraction from abdominal recordings using multi-scale principal components analysis," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 6507–6510.

Hamidreza Ghonchi holds an M.Sc. degree in artificial intelligence, and graduated from the Shahrood University of Technology, Iran in 2019. His research interests are in bio-signal processing, machine learning, and deep learning.

Vahid Abolghasemi (M'12-SM'17) is currently an Assistant Professor at the School of Computer Science and Electronic Engineering, University of Essex, UK. He received his PhD degree from the University of Surrey, Guildford, the UK in 2011. His main research interests include signal and image processing, compressive sensing, and machine learning.