Deep Learning for Detection of Fetal ECG from Multi-Channel Abdominal Leads

Fang-Wen Lo and Pei-Yun Tsai
Department of Electrical Engineering,
National Central University, Taiwan

***Abstract*—In this paper, we propose to use a CNN-based approach for fetal ECG detection from the abdominal ECG recording. Our work flow contains a pre-processing phase and a classification phase. In the pre-processing phase, abdominal ECG waveform is normalized and segmented. Then, short-time Fourier transform is applied to obtain time-frequency representation. The 2D representation is sent to 2D convolutional neural network for classification. Two convolutional layers, two pooling layers, one fully-connected layer are used. The softmax activation function is used at the output layer to compute the probabilities of four events. The classified results from multiple channels are fused to derive the final detection according to the respective detection accuracies. Compared to the K-nearest neighbor algorithm, the CNN-based classifier has better detection accuracy.**

***Index Terms*—Electrocardiogram (ECG), fetal ECG, abdominal ECG, convolutional neural network, classification.**

I. Introduction

Abdominal electrocardiogram recording provides helpful information for evaluating the health condition of the fetus during pregnancy. To avoid the risk of infection and injury, the non-invasive approach is preferred to the invasive one. The surface electrodes are placed on the mother's abdomen in the non-invasive approach and the fetal electrocardiogram (fECG) can be monitored. However, the measured signal contains various types of disturbances, including the maternal electrocardiogram (mECG), noise, maternal muscle and respiration activity as well as fetal movement. Most of the undesired disturbances do not overlap the fECG in the similar frequency band and thus can be removed by the conventional filtering or frequency-domain processing except the mECG, which has larger amplitude and similar frequency as fECG [1]. During these years, various signal processing techniques are developed to extract or detect the fECG so as to enhance the reliability of recording for accelerating its applications in clinical diagnosis.

Blind source separation (BSS) has been shown to outperform the matched filtering and correlation techniques for fECG extraction [1]. Independent component analysis (ICA), one of the algorithms for BSS, has been widely investigated [2]-[5]. The fast ICA algorithm is employed to multi-channel recording for fECG extraction [3][4]. In [4], besides fast ICA, post-processing is performed, which uses undecimated wavelet transform and fast Fourier transform as well as inverse Fourier transform to denoise so that the fetal heart rate can be detected accurately. In [5], ICA and compressed sensing are combined. Then, ℓ_p-regularized recursive least squares algorithm is used for reconstruction. One the other hand, there are some works which partition the electrocardiogram into waveform segments and detect the event in each segment [6][7]. In [6], the waveform segments are classified into F-event that has fetal QRS (fQRS) complex only, M-event that has maternal QRS (mQRS) complex only, MF-event in which both fQRS and mQRS present, and N-event that contains only noise. Machine learning algorithms, such as support vector machine (SVM), K-nearest neighbor (KNN), and Bayesian network (BN) are often adopted. Recently, due to the success of deep learning for image recognition and classification, one-dimensional convolutional neural network is used in [7], and events of ventricular ectopic beats and supraventricular ectopic beats are detected with the time-domain ECG features. In [8], extended Kalman smoother and template adaption are used for fetal heart rate detection. Besides, time-frequency analysis is suggested in [9]-[11] for ECG feature extraction. Wigner–Ville distribution (WVD) is used in [9] while short time Fourier transform (STFT) can be seen in [10]-[12]. Time-frequency representation yields good detection performance [11][12].

In this paper, we aim to use the deep learning technique to deal with the fetal ECG detection. In light of the above, multi-channel abdominal ECG signals are processed by the short-time Fourier transform (STFT) to obtain the time-frequency representation so that two-dimensional features can be extracted. Consequently, deep neural network is adopted which contains two convolutional layers and one fully-connected layer. Because of multi-channel leads, a fusion stage is inserted to derive the final detection result. The abdominal ECG signals in the PhysioNet database [13] are used. Better detection performance is achieved by using the proposed deep learning scheme than using the K-nearest neighbor method.

In the following, Sec. II first introduces the methodology for multi-channel abdominal ECG detection. The two-dimensional convolution neural network for feature extraction is illustrated in Sec. III. In Sec. IV, simulation results are given to show the advantages of the proposed scheme. Finally is the conclusion.

II. System Description and Methodology

The multi-channel abdominal ECG recording from PhysioNet [13] is used. Each recording consists of four signals measured from maternal abdomen and one reference signals from fetal head. The sampling rate (f_s) is 1KHz with 16-bit resolution for 5-minute recording. Fig. 1 describes the flow of our proposed classifier. Two phases are included, namely the preprocessing phase and the classification phase.

In the preprocessing phase, the ECG signal is first normalized to keep the value in the range of $[-1,1]$, and the normalized output is given by

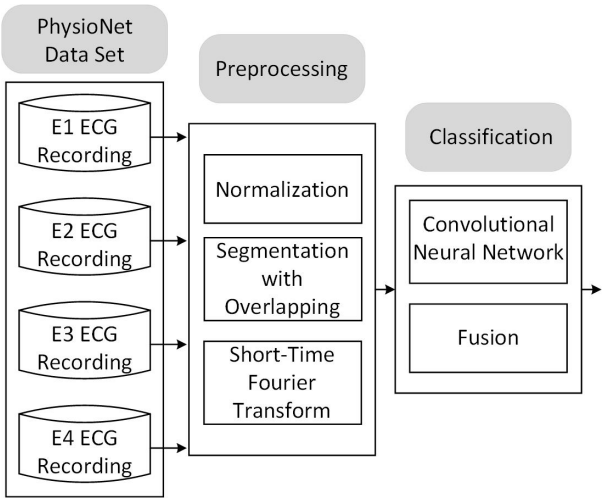


Fig. 1 Workflow of the abdominal ECG classifier.

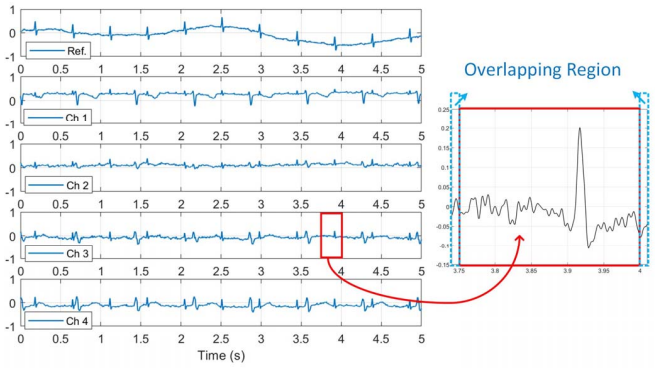


Fig. 2 Segmentation of the abdominal ECG signals and the reference fetal ECG.

$$\bar{x}_{i,n} = 2\frac{x_{i,n}-x_{i,min}}{x_{i,max}-x_{i,min}} - 1, \quad (1)$$

where $x_{i,n}$ is the signal from the ith lead, $x_{i,min}$ and $x_{i,max}$ is the minimum and maximum of $x_{i,n}$ for $0 \le n \le N$ and $\frac{N}{f_s} = 120s$. Then, $\bar{x}_{i,n}$ is segmented and each segment has 250 ms plus 10-ms overlapping region at two ends to avoid the fragmentation of the QRS complex near the 250-ms boundary as shown in Fig. 2. The top sub-figure is the reference fetal ECG and the four bottom sub-figures are the abdominal ECG of four leads.

The two-dimensional (2D) time-frequency representation is obtained by short-time Fourier transform (STFT), which is described by

$$X_i(n,k) = \sum_{m=0}^{M-1} \bar{x}_i(m+n\Delta)w(m)e^{-j2\pi km/M}, \quad (2)$$

where Δ is the time step or the hop size and is set to 6; $w(m)$ is the Hamming window described by

$$w(m) = \begin{cases} 0.53836 - 0.46164\cos\left(\frac{2\pi m}{\widetilde{M}-1}\right) & 0 \le m \le \widetilde{M}-1, \\ 0 & else \end{cases} \quad (3)$$

and the window size $\widetilde{M}$ is 36. The 256-point FFT is applied. Consequently, the dimension of the time-frequency representation is 40×128. For the 120s waveform, 480 plots are generated for each lead. Thus, the a total of 1920 2D time-frequency representations are then grouped and labelled as the M-event, F-event, MF-event, and N-event as in [6]. The spectrograms for the respective events are shown in Fig. 3. It is clear that the spectrogram of the N-event is different from the other three.

The 1920 time-frequency plots are fed into the convolutional neural network (CNN), some for training and some for testing. Besides, for each index n, the waveform segments from four leads should be corresponding to one single event. Thus, fusion of the four detection results is considered. The details about the classification will be addressed in the next section.

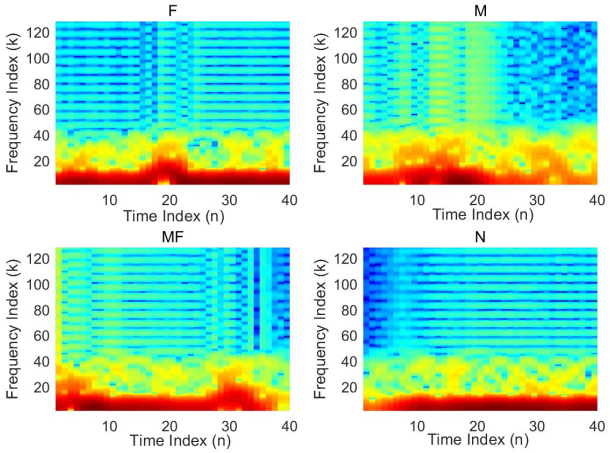


Fig. 3 Spectrogram of M-event, F-event, MF-event, and N-event.

III. Convolutional Neural Network for Abdominal ECG Classification

Since time-frequency representation has been shown to achieve good detection results [11][12], we then use 2D CNN to deal with the time-frequency features for abdominal ECG classification. Fig. 4 shows the entire architecture of the 2D CNN. It contains two convolutional layers, two pooling layers and one fully-connected layer. In the following, parameters and hyperparameters of each layer will be described. Without loss of generality, subscript i to indicate the lead index is dropped here.

A. Forward propagation

The kernel size of the first convolutional layer is 17×17 and six neurons exist in the layer. During forward propagation, given the 40×128 input, the output feature map with the stride equal to 1 has the size of 24×112, which can be described by

$$\mathbf{Y}_p^{(l)} = \mathrm{b}_p^{(l)} + \sum_{p'=0}^{P_{l-1}-1} 2D\_Conv(\mathbf{W}_{p,p'}^{(l-1)}, \mathbf{Z}_{p'}^{(l-1)}), \quad (4)$$

where $\mathbf{Y}_p^{(l)}$, $b_p^{(l)}$, and $\mathbf{Z}_{p'}^{(l-1)}$ are the output of 2D convolution, the bias of the pth neuron at layer l and the output from layer $l-1$, respectively; $\mathbf{W}_{p,p'}^{(l-1)}$ is the weight from neuron p' at layer $l-1$ to neuron p at layer l; P_{l-1} is the number of neurons at layer $l-1$. The sigmoid defined by

$$\sigma(y) = \frac{1}{1+e^{-y}} \quad (5)$$

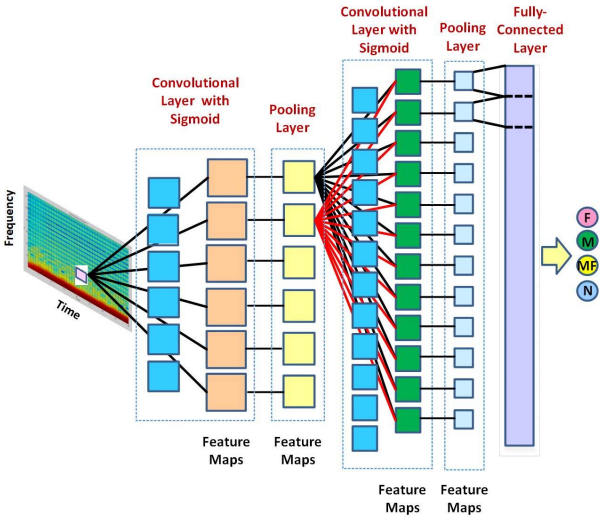


Fig. 4 Convolutional neural network architecture for abdominal ECG classification.

is used as the activation function, which squashes the input range $[-\infty, +\infty]$ into the range $[0,1]$. Thus,

$$\mathbf{Z}_p^{(l)}(u,v) = \sigma\left(\mathbf{Y}_p^{(l)}(u,v)\right), \quad \forall\, u, v. \tag{6}$$

The pooling layer follows the convolutional layer and average pooling is used. The operation at the polling layer takes the form of

$$\mathbf{Z}_p^{(l)}(c,r) = \frac{1}{4}\sum_{\alpha=0}^{1}\sum_{\beta=0}^{1}\mathbf{Z}_p^{(l-1)}(2c+\alpha, 2r+\beta). \tag{7}$$

The dimension of the feature map becomes 12×56 at the output of the first pooling layer. Then, the second convolutional layer contains 12 filters, each having the kernel size of 5×5. After applying Eqs. (4), (6), and (7) for the second 2D convolution and average pooling, the size of the output feature map becomes 4×26. A fully-connected layer is inserted after the second pooling layer. Therefore, the output structure of the second pooling layer is flattened to create a 1-D feature vector. In the fully-connected output layer, the pth neuron generates the scalar output, which is given by

$$\mathrm{Y}_p^{(L)} = \mathrm{b}_p^{(L)} + \sum_{p'=0}^{P_{L-1}-1} \mathrm{w}_{p,p'}^{(L-1)}\, \mathrm{Z}_{p'}^{(L-1)}, \tag{8}$$

where $\mathrm{Z}_{p'}^{(L-1)}$ is the p'th element in the flattened feature vector and $P_{L-1} = 1248$. The softmax activation function is adopted in the output layer L,

$$\mathrm{Z}_p^{(L)} = \sigma\left(\mathrm{Y}_p^{(L)}\right) = \frac{e^{\mathrm{Y}_p^{(L)}}}{\sum_{k=0}^{P-1} e^{\mathrm{Y}_k^{(L)}}}. \tag{9}$$

Thus, any real value of the the pth neuron is squashed into the range $[0,1]$ to denote the probability of the pth class.

B. Back propagation

Denote the target class vector as $\mathbf{t} = [t_0 \quad \cdots \quad t_{P_L-1}]$, where $t_p \in \{0,1\}$ and $P_L = 4$. The cross entropy loss function is used and is given by

$$\xi = -\sum_{p=0}^{P_L-1} t_p ln(\mathrm{Z}_p^{(L)}). \tag{10}$$

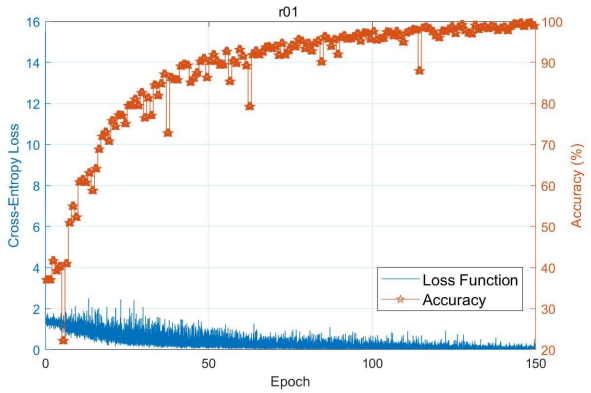


Fig. 5 Loss function versus detection rate during training phase.

During back propagation, the sensitivity of the loss function to $\mathrm{Y}_p^{(L)}$, the input of the softmax function, can be expressed as

$$\frac{\partial \xi}{\partial \mathrm{Y}_p^{(L)}} = \frac{\partial \xi}{\partial \mathrm{Z}_p^{(L)}}\frac{\partial \mathrm{Z}_p^{(L)}}{\partial \mathrm{Y}_p^{(L)}} = \left(\mathrm{Z}_p^{(L)} - t_p\right). \tag{11}$$

And the sensitivity to $\mathrm{Z}_p^{(L-1)}$, the output of the fully-connected layer, is given by

$$\frac{\partial \xi}{\partial \mathrm{Z}_p^{(L-1)}} = \sum_{p'=0}^{P_L-1}\frac{\partial \xi}{\partial \mathrm{Y}_{p'}^{(L)}}\frac{\partial \mathrm{Y}_{p'}^{(L)}}{\partial \mathrm{Z}_p^{(L-1)}} = \sum_{p'=0}^{P_L-1}\left(\mathrm{Z}_{p'}^{(L)} - t_{p'}\right)\mathrm{w}_{p',p}^{(L-1)}. \tag{12}$$

For average pooling defined by (6), the sensitivity to the input feature map is

$$\frac{\partial \xi}{\partial \mathbf{Z}_p^{(l-1)}(u,v)} = \frac{1}{4}\frac{\partial \xi}{\partial \mathbf{Z}_p^{(l)}(\lfloor \frac{u}{2} \rfloor, \lfloor \frac{v}{2} \rfloor)}, \tag{13}$$

and actually upsampling is performed. The derivative of the sigmoid function is written as

$$\sigma'(y) = \sigma(y)\left(1 - \sigma(y)\right). \tag{14}$$

Consequently, the sensitivity of the loss function to the input of the convolutional layer $\mathbf{Z}_p^{(l)}$ described in (4) becomes

$$\frac{\partial \xi}{\partial \mathbf{Z}_p^{(l-1)}(u,v)} = \sum_{p'=0}^{P_l-1}\sum_{u'}\sum_{v'}\Delta_{p'}^{(l)}(u',v')\frac{\partial \mathbf{Y}_{p'}^{(l)}(u',v')}{\partial \mathbf{Z}_p^{(l-1)}(u,v)}$$
$$= \sum_{p'=0}^{P_l-1} 2D\_conv\_z\left(\Delta_{p'}^{(l)}(u',v'), Rot(\mathbf{W}_{p',p}^{(l-1)})\right), \tag{15}$$

where

$$\Delta_{p'}^{(l)}(u',v') = \frac{\partial \xi}{\partial \mathbf{Z}_{p'}^{(l)}(u',v')}\sigma\left(\mathbf{Y}_{p'}^{(l)}(u',v')\right)\left(1 - \sigma\left(\mathbf{Y}_{p'}^{(l)}(u',v')\right)\right); \tag{16}$$

$2D\_conv\_z(,)$ is the 2D convolution with zero padding; $Rot(\mathbf{W}_{p',p}^{(l-1)})$ is to rotate the kernel by 180°.

Fig. 5 shows the loss function and the detection accuracy of the CNN-based classifier for the abdominal ECG recording r01 during the training phase. The batch size is 8 and the waveform segments of E1 to E4 are randomly distributed in one epoch, which has 108 batches in total. It is clear that the kernel is

effectively adapted to converge and the detection accuracy is also improved gradually.

Table I Confusion metrics before and after fusion

E1 Classification Result				
Ground Truth	F	M	MF	N
F	37	2	5	1
M	8	17	13	1
MF	0	11	27	0
N	18	0	0	64
E2 Classification Result				
Ground Truth	F	M	MF	N
F	51	10	16	14
M	7	17	12	2
MF	3	0	17	0
N	2	3	0	50
E3 Classification Result				
Ground Truth	F	M	MF	N
F	51	5	3	17
M	4	21	2	0
MF	6	4	40	0
N	2	0	0	49
E4 Classification Result				
Ground Truth	F	M	MF	N
F	63	1	3	3
M	0	27	7	0
MF	0	1	35	0
N	0	1	0	63
Fusion Classification Result				
Ground Truth	F	M	MF	N
F	62	2	5	2
M	0	27	4	0
MF	0	1	36	0
N	1	0	0	64

C. Fusion

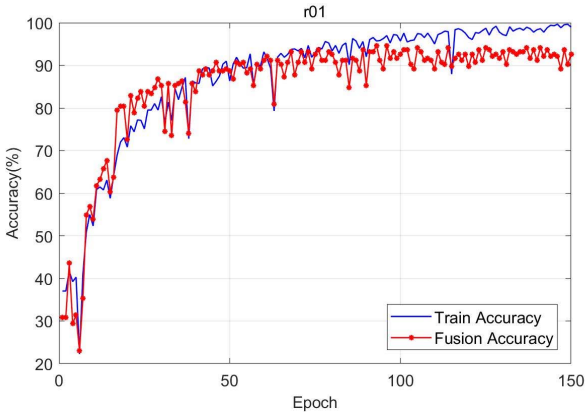
The recording from some leads may suffer noise or some disturbance and thus waveform segments from different leads at the same time may be detected as different classes. A fusion is required to generate one final result. Given $\mathbf{z_i} = [Z_{i,0}^{(L)} \quad \dots \quad Z_{i,P_L-1}^{(L)}]^T$, where $Z_{i,p}^{(L)}$ denotes the softmax output of the pth neuron for the waveform segment of the ith lead, the soft fusion output is computed by

$$\boldsymbol{s} = \frac{\sum_{i=1}^{4} g_i \mathbf{z_i}}{\sum_{i=1}^{4} g_i}. \tag{17}$$

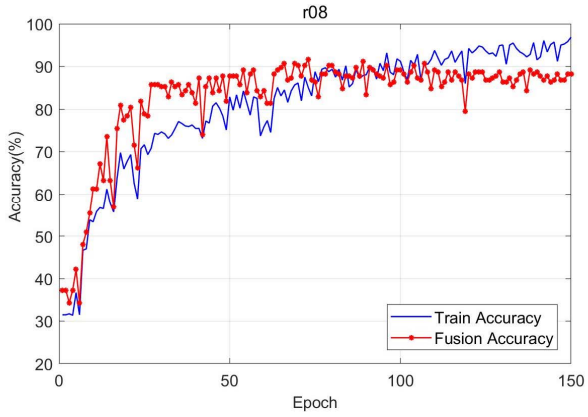
The fusion weight g_i is obtained by the detection accuracy of extra 240 waveform segments, 60 segments from each lead. The confusion metrics of recording r01 before and after the fusion is given in Table I. The detection accuracy is raised to 92.65% after fusion.

IV. Simulation and Comparison

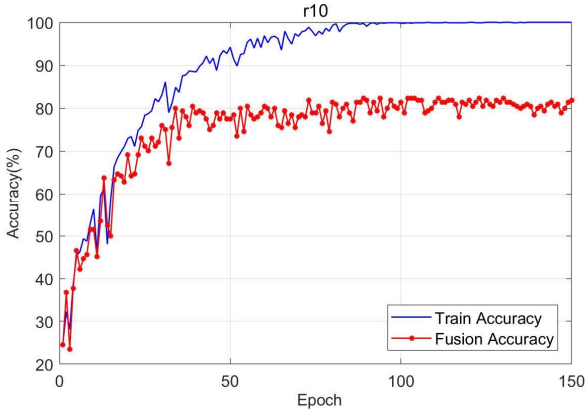
To verify the performance of the proposed 2D CNN classifier, more simulations are conducted. In the following, 120s ECG recording of four channels is used, which can generate 1920 waveform segments. We use 864 segments for training, 240 segments to compute fusion weight g_i, and 816 segments for inference. After fusion, 204 results are obtained. In Fig. 6, we show the training accuracy and inference accuracy after fusion versus the number of epoch for training. It is clear that for different ECG recording signals, 80% to 95% inference accuracy can be achieved. If we use the mixed waveform



(a)



(b)



(c)

Fig. 6 The training accuracy versus the inference accuracy after fusion for (a) r01 recording, (b) r08 recording, and (c) r10 recording.

segments from three pregnant women to train the 2D CNN, the result is given in Fig. 7. In this case, there are 2592 training segments. The fusion weights are derived from the respective pregnant women and the average inference accuracy is depicted. We can that see a larger training size bring better average inference accuracy, which is around 90%.

The KNN algorithm is adopted for comparison. We follow the procedure in [6] including segmentation, normalization,

applying Haar-wavelet transform, principle component analysis, and sampling without replacement techniques. Note that the missing sample replacement mentioned in [6] is not adopted for the fair comparison with CNN. Table II shows the comparison of the CNN classifier and the KNN classifier. The inference accuracy of the CNN is obtained for training 150 epochs. The best detection accuracy of KNN among the 10-fold cross validation is given. Although the frequency-domain features of the abdominal ECG are utilized by two classifiers, the CNN algorithm outperforms the KNN algorithm.

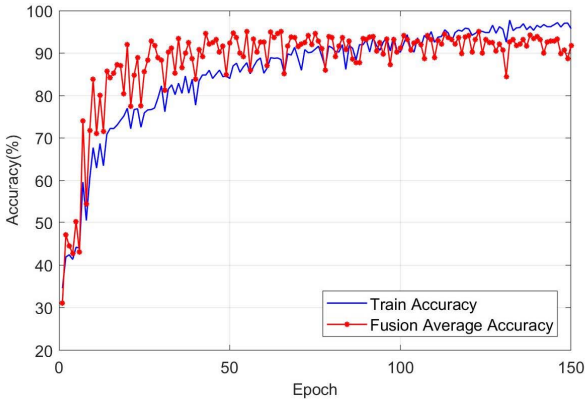


Fig. 7 The average detection accuracy given training the CNN with mixed waveforms from three pregnant women.

Table II Detection accuracy comparison of CNN and KNN.

Accuracy	r01	r08	r10
CNN	92.65%	88.24%	81.86%
KNN	83.33%	79.17%	70.83%

V. CONCLUSIONS

A 2D CNN is used for abdominal ECG classification. The short-time Fourier transform is applied to generate 2D time-frequency representation as the input to the 2D CNN so that the time-frequency features can be extracted. Two convolutional layers, two pooling layers and one fully-connected layers are used in the neural network. Sigmoid activation function and softmax activation function are used in the convolutional layers and in the fully-connected layer, respectively. After four detection results of the respective channels are obtained, fusion by weighted average of the softmax outputs is performed to generate the final detection. From the simulation result, the CNN classifier has better performance than the KNN classifier for the fetal ECG detection.

REFERENCES

[1] S. Ravindrakumar and K. Bommanna Raja, "Fetal ECG Extraction and Enhancement in Prenatal Monitoring – Review and Implementation Issue," Trendz in Information Sciences & Computing (TISC2010), 2010, pp. 16-20.

[2] H. Wan, Q. Liu and J.Chai, "A method for extracting fECG based on ICA algorithm", 9th IEEE International Conference on Signal Processing, Beijing, China, pp.2761–2764, 26-29 October 2008.

[3] S. Nikam and S.Deosarkar, "Fast ICA Based Technique for non-invasive fetal ECG extraction," 2016 Conference on Advances in Signal Processing (CASP), 2016. Pp. 60-65.

[4] V. Ionescu, M. Hnatiuc, "Fetal Heart Rate Detection and Monitoring From noninvasive Abdominal ECG Recordings, " IEEE International Conference on E-Health and Bio-Engineering, 2015, pp. 1-4.

[5] D. Gurve, J, K. Pant, S. Krishnan, "Real-time fetal ECG extraction from multichannel abdominal ECG using compressive sensing and ICA," Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 2794-2797.

[6] J. A. Delgado, M. Altuve, M. N. Homsi, "Haar wavelet transform and principal component anaylsis for fetal QRS classfification from abdominal maternal ECG recordings," Symposium on Signal Processing, Images and Computer Vision, 2015, pp. 1-6.

[7] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG by 1-D convolutional neural networks," IEEE Transactions on Biomedical Engineering, pp. 664-675, Mar. 2016.

[8] F. Andreotti and et al., "Robust fetal ECG extraction and detection from abdominal leads," Physiol Meas., pp.1551-1567, Aug. 2014

[9] E. C. Karvounis, M. G. Tsipouras, D. I. Fotiadis, K. K. Naka, "An automated methodology for fetal heart rate extraction from the abdominal electrocardiogram," IEEE Transactions on Information Technology in Biomedicine, pp. 628-638, Jun. 2006.

[10] B. T. Krishna, "Fetal ECG extraction using time-frequency analysis techniques," International Conference on Robotics and Automation Sciences, 2017, pp. 167-171.

[11] S. Y. Chun, J. H. Kang, H. Kim, C. Lee, I. Oakley, S. P. Kim, "ECG based user authentication for wearable devices using short time Fourier transform," International Conference on Telecommunications and Signal Processing (TSP), pp. 656-659, 2016.

[12] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbaugh, "ECG biometrics: A robust short-time frequency analysis," in 2010 IEEE International Workshop on Information Forensics and Security (WIFS), Oct. 2010, pp. 1–6.

[13] L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P. Ivanov R.M. R.G. Mark , J.E. Mietus , G.B. Moody , C-K Peng and H. Stanley , "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals" , Circulation, 101(23), pp.215-220, 13 June 2000.