2021 13th International Conference on Wireless Communications and Signal Processing (WCSP) | 978-1-6654-0785-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/WCSP52459.2021.9613207

Machine-Learning-based Aortic Stenosis Detection for Electronic Stethoscope

Zhen Shi
School of Electronics and Information Engineering
Soochow University
Suzhou, China
20195228026@stu.suda.edu.cn

Neng Dai
Department of Cardiology
Zhongshan Hospital, Fudan University
National Clinical Research Center for Interventional Medicine
Shanghai, China
niceday1987@hotmail.com

Renyu Liu
School of Electronics and Information Engineering
Soochow University
Suzhou, China
1002280496@qq.com

Jaijun Wang
School of Electronics and Information Engineering
Soochow University
Suzhou, China
jjwang@suda.edu.cn

Shengsheng Cai
Suzhou Melodicare Medical Technology Co., Ltd.
Suzhou, China
caishengsheng@mintti.cn

Nan Hu
School of Electronics and Information Engineering
Soochow University
Suzhou, China
hunan@suda.eu.cn (corresponding author)

***Abstract*—Analyzing the phonocardiogram (PCG) collected by a electronic stethoscope can help to quickly diagnose structural heart diseases. In this paper, we aim at automatic detection of aortic stenosis (AS) based on PCG, with the aid of two proposed machine-learning based methods. The first method is a model-dependent method, a Gaussian mixture model – hidden Markov model (GMM-HMM) method, which exploits the temporal relationship among states in cardiac sounds. The second one is a data-dependent method, implemented by a 1D/2D-fused-feature-based convolutional neural network. The results of comparative experiments showed that both methods can fulfill the AS automatic diagnosis task to a certain extent, and in most cases CNN scored higher than GMM-HMM, which indicated the importance of automatically learning an unknown model from data in this problem, although the GMM-HMM method with fewer parameters also have potential advantages in practice.**

Keywords—phonocardiogram, aortic stenosis, Gaussian mixture model – hidden Markov model, convolutional neural network

I. INTRODUCTION

Electronic stethoscope is a portable and low-cost device that can quickly acquire cardiopulmonary sounds, where phonocardiogram (PCG) [1] can be used for diagnosis of heart diseases, especially for structural heart diseases. Automatic recognition of pathological PCGs caused by structural heart diseases corresponds to applying combinations of features and classifiers in classification tasks. At present, the automatic analysis of PCG is mainly achieved through two methodologies: traditional pattern recognition methods or artificial-neural-network-based features extraction and classification.

A simple way is to extract the temporal or spectral characteristics of PCG, and then use empirical thresholds for classification. However, the subjectively selected features or thresholds may not work well confronting individual difference. Support vector machine (SVM), a non-probabilistic binary linear or nonlinear machine learning method has been used [2-5], where various kernels were involved. The role of kernel functions playing in improving training efficiency and achieving automatic PCG classification was discussed in [2]. Choi *et al.* performed wavelet-packet-based PCG decomposition, calculated the energy distribution, and realized classification by SVM [3]. Al-Naami *et al.* [4] derived combination of multiple features in the time-frequency domain, and used SVM for detection of the paradoxical splitting in the second heart sound. Zhang *et al.* [5] used binary tree SVM (BT-SVM) as a classifier for distinguishing mitral stenosis (MS), ventricular septal defect (VSD), and aortic stenosis (AS). Another widely used classifier is k-Nearest Neighboors (k-NN), *e.g.* Bentley *et al.* [6] derived a variety of features as the input into a k-NN-based heart murmur detector.

Due to the self-organizing, real-time, and self-adaptive learning characteristics, neural networks have also been used in cardiac abnormality detection. Reed *et al.* derived features of the PCG segments using 7-layer wavelet decomposition with 4th-order Coiflet wavelet basis function, and built a 3-layer neural network for classification [7]. Higuchi *et al.* also established a 3-layer neural network as a classifier for identifying 9 typical pathological cardiac sounds [8]. Tschannen *et al.* used a neural network for feature extraction instead, and formed the classifier by SVM [9]. Potes *et al.* proposed a joint classification method based on AdaBoost and convolutional neural networks (CNN), to distinguish normal cardiac sounds from the abnormal ones [10]. Maknickas *et al.* took the Mel-frequency spectral coefficients (MFSC) of the cardiac sounds as the input feature of CNN [11]. Chen *et al.* used K-Means to refine the Mel-frequency cepstral coefficients (MFCCs) features, and then input the refined features into a deep neural network (DNN) for classification [12]. Rouhani *et al.* used independent component analysis (ICA) to extract 32 features, picked out four most informative ones, and compared the performances of SVM and neural network on identifying pathological cardiac sounds [13]. Thomae *et al.* proposed an end-to-end method, which directly took the raw cardiac sound as input, and extracted features and made classification by 1-D CNN and recurrent neural network (RNN) [14].

In the existing methods, the PCG classification tasks were usually very superficial, *e. g.* simply discriminating normal cardiac sounds from abnormal ones, while the abnormality usually did not include any pathological information. Several studies concerned on classifying the cardiac sounds into some classes corresponding to different

This work was supported by Suzhou Science and Technology Project under Grant SYS2019029.

978-1-6654-0785-4/21/$31.00 ©2021 IEEE

structural heart diseases, while the number of samples in each class was usually very limited. In this paper, we focus on the problem of detecting AS from normal/abnormal mixed cardiac sounds. We used 199 AS cardiac sound recordings provided by cooperative hospitals, and 221 normal cardiac sound recordings. We study two machine-learning-based AS detection method, where the first one is model-dependent realized by Gaussian mixture model – hidden Markov model (GMM-HMM), and the second one is data-dependent given by a established 1D/2D-fused-feature-based CNN. The performances of AS detection by the two proposed methods are explored and compared, to reveal the advantages of these two methods in their most applicable scenarios.

II. PCG of AS Patients

A. Data Collection

In this study, AS PCG recordings were given by our cooperative hospitals, providing total 265 recordings in their routine medical examinations using the electronic stethoscope (Smartho-D2) developed by *Melodicare*. After assessing the recording quality by specialists, 199 recordings were reserved in this study. In addition, 221 normal PCG recordings by healthy volunteers were also included. The collected PCG recordings all had more than one systole-diastole cycle, and they were partitioned into segments with each one containing an intact course of first heart sound (S1)-systole-second heart sound (S2). We got 3303 normal cardiac sound segments and 4070 AS cardiac sound segments at last. To train and test a machine-learning method for cardiac sound classification, the data collection was further partitioned into two data sets: the training set and the test set. The training set contained 2655 normal segments and 3243 AS segments. The test set had 648 normal segments and 827 AS segments. We guaranteed that a cardiac sound segment in the training set and the one in the test set did not belong to a same PCG recording.

B. A Glance at the AS Cardiac Sound Segment

AS may induce pathological murmurs during the systolic phase of heart sounds compared to the normal PCG, and severe AS may even cause the main components of heart sounds to disappear. Fig. 1 shows examples of cardiac sound segments of normal subject, mild AS, moderate AS, and severe AS. It can be found that as the symptom goes severe, the systolic murmurs may become more and more obvious. According to the observed phenomenon, we design machine learning methods to recognize state transitions or special patterns in a cardiac sound segment, hence realizing AS detection.

III. GMM-HMM-based AS Detection

GMM is a widely used machine learning model due to its simple formulation and calculation convenience. Its distribution is a composition of multiple Gaussian distributions, which is formulated as

$$p(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k \phi\left(\mathbf{x} \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad (1)$$

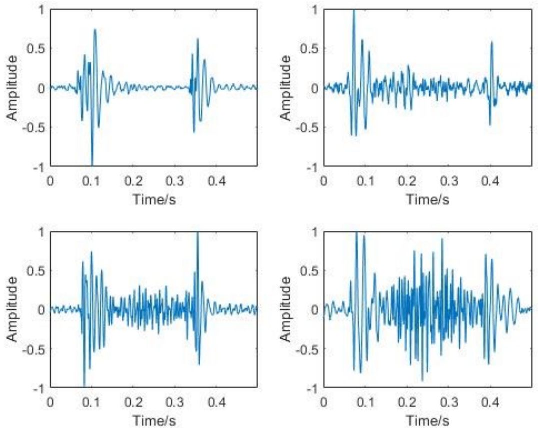

Fig. 1. Examples of PCG segments of healthy subject or AS patients. Top left: healthy subject; top right: mild AS; left bottom: medium AS; right bottom: severe AS.

where K denotes the number of Gaussian distributions, α_k is the weight of the k th Gaussian distribution, $\alpha_k \geq 0$ and $\sum_{k=1}^{K} \alpha_k = 1$, $\phi\left(\mathbf{x} \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$ is the probability density function (PDF) of the k th Gaussian distribution, given by

$$\phi\left(\mathbf{x} \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = (2\pi)^{-D/2} \left|\boldsymbol{\Sigma}_k\right|^{-1} e^{-(\mathbf{x}-\boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}-\boldsymbol{\mu}_k)/2}, \quad (2)$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the mean vector and covariance matrix of the k th Gaussian distribution, respectively, D is the dimension of $\mathbf{x}$, and $(\cdot)^{\mathrm{T}}$ denotes matrix transpose.

The parameters training of GMM is given by an expectation-maximization (EM) process, depicted as:

(1) E-step: if there is a data collection $\{\mathbf{x}_j\}_{j=1}^{n}$ with each data sample obeying GMM independently and the class of Gaussian distribution corresponding to $\mathbf{x}_j$ is denoted as z_j , the posterior probability of $z_j = k$ is given by

$$\gamma_{jk} = \frac{\phi\left(\mathbf{x} \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{k=1}^{K} \alpha_k \cdot \phi\left(\mathbf{x} \middle| \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}, \quad k = 1, 2, ..., K , \quad (3)$$

where $\gamma_{jk} = P\left(z_j = k \middle| \mathbf{x}_j\right)$.

(2) M-step: we update mean vectors, covariance matrices and weights of K classes of Gaussian distributions according to γ_{jk} calculated by E-step:

$$\boldsymbol{\mu}_k^{\mathrm{new}} = \sum_{j=1}^{n} \gamma_{jk} \mathbf{x}_j / \sum_{j=1}^{n} \gamma_{jk} , \quad (4)$$

$$\boldsymbol{\Sigma}_k^{\mathrm{new}} = \sum_{j=1}^{n} \gamma_{jk} \left(\mathbf{x}_j - \boldsymbol{\mu}_k^{\mathrm{new}}\right)\left(\mathbf{x}_j - \boldsymbol{\mu}_k^{\mathrm{new}}\right)^{\mathrm{T}} / \sum_{j=1}^{n} \gamma_{jk} , \quad (5)$$

$$\alpha_k^{\mathrm{new}} = \sum_{j=1}^{n} \gamma_{jk} / n . \quad (6)$$

The E-step and the M-step go alternatively, until convergence.

HMM can be denoted as a 3-element model $\lambda = \{\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}\}$, where $\boldsymbol{\pi}$ represents the initial probability distribution vector, $\mathbf{A}$ represents the state transition probability matrix, and $\mathbf{B}$ represents the observation probability matrix. According to the state transition style, HMM can be categorized into ergodic HMM, left-to-right HMM, etc. Since the state transition between cardiac sound components has a fixed pattern, as shown in Fig.2, so left-to-right HMM is used in this paper.

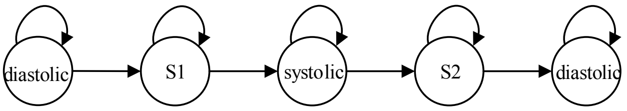

Fig. 2. Cardiac sound state transition topology diagram

GMM-HMM combines GMM and HMM, and uses the probability generated by GMM as the observation probability of the observed data in HMM in a certain state. In this paper, the MFCCs of cardiac sound segments were extracted as observations. There are 4 states in the cardiac sound signal: S1, systole, S2, and diastole. As we have partitioned the cardiac sound recordings into segments containing S1, systole, and S2, and the start and end of each segment are diastolic, we chose 5 as the number of states in HMM. Fig. 3 shows the GMM-HMM used in this study.

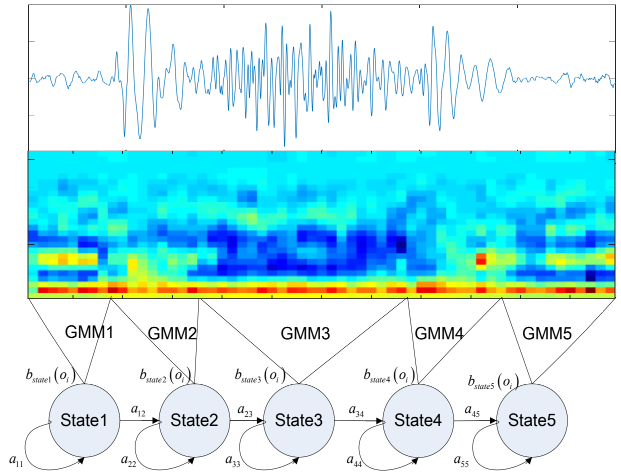

Fig. 3. GMM-HMM used for cardiac sound classification.

Fig. 4 shows a schematic diagram of training and recognition. Fig. 4(a) shows that, during training, a model M_1 and a model M_2 were trained using normal cardiac sound data and AS cardiac sound data, respectively. Fig. 4(b) is a schematic diagram of recognition. During recognition, the posterior probabilities $P(O|M_1)$ and $P(O|M_2)$ belonging to the normal model M_1 and the AS model M_2, respectively, were calculated for the input data, and the largest probability corresponds to the recognized class.

IV. 1D/2D-Fused-Feature-based CNN

In establishing a neural network for classification task, to solve the vanishing gradient problem, making a wider network is one of the favorable choices. In this paper, we built a wider network by fusing 1-D feature and 2-D feature of a cardiac sound segment.

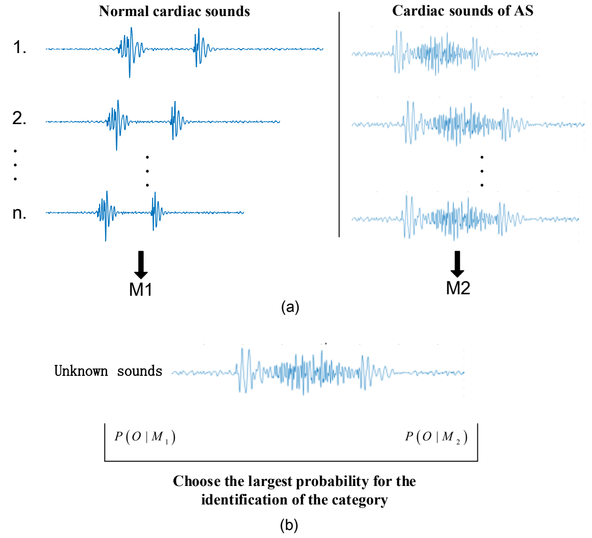

Fig. 4. Schematic diagram of training and recognition in GMM-HMM.

The 2-D feature was obtained by utilizing the MFCCs of the signal. Considering that MFCCs characterize the static features of cardiac sound signals, in order to address more useful information, we additionally extracted 1st-order and 2nd-order differences of MFCCs to evaluate the dynamic information, thus forming a 3-channel MFCC-based input tensor. This 3-channel tensor was input to a 2-D CNN to extract the corresponding 2-D feature. As some information may be lost when only using MFCCs, in order to retain more useful information, the 1-D cardiac sound signals was input to a parallel 1-D CNN to give 1-D feature. The 1-D feature and 2-D feature were concatenated to form a fused feature, based on which the classification task was carried out.

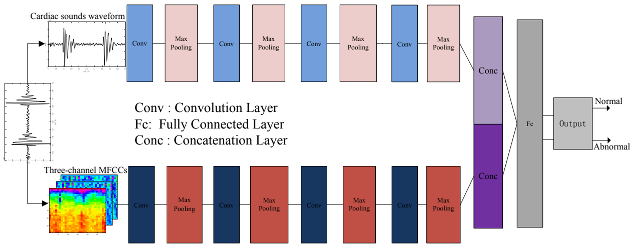

Fig. 5. The structure of the proposed neural network for AS detection.

The structure of the proposed neural network is displayed in Fig. 5. The 1-D cardiac sound signal is sent to two pathways after the original input. One utilizes direct input, and features are extracted through four 1-D convolution-pooling modules. The other one first derives 3-channel MFCC-based tensor, and then uses four 2-D convolution-pooling modules for feature extraction. The features extracted by the 2-D CNN are spread into one-dimension and concatenated with the output feature of 1-D CNN. The importance of 1-D feature or 2-D feature is evaluated by different weights, which will be explored in experiments to derive the optimal values. By passing the concatenated features through a fully connected layer and Softmax activation function, we ultimately obtain two probabilities, where the larger one indicates the winner in classification. In the 1-D convolution-pooling modules, kernel sizes of 1-D convolution are 300, 270, 240, and 210 in sequence, and strides of convolution and maximum pooling are 2 and 20, respectively. In the 2-D convolution-pooling modules, kernel sizes of convolution layers and maximum pooling layers are all 3×3, and strides of them are all 2×2. In the training procedure, the optimizer was Adam and regularization and dropout were added to prevent overfitting. The network training was implemented by Keras, on a workstation equipped with 2 NVIDIA GeForce GTX 1080 Ti GPUs.

V. Performance Evaluation

The partition of data collection for training and testing has been depicted in Section II, for both GMM-HMM and 1D/2D-fused-feature-based method. To evaluate the performances of AS detection, 5 indicators were used: accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), and F1-score (F_1). These indicators were defined as follows:

$$\mathrm{ACC} = (TN + TP)/(TN + TP + FP + FN), \tag{7}$$

$$\mathrm{PRE} = TP/(TP + FP), \tag{8}$$

$$\mathrm{SEN} = TP/(TP + FN), \tag{9}$$

$$\mathrm{SPE} = TN/(TN + FP), \tag{10}$$

$$\mathrm{F}_1 = 2 \times \mathrm{PRE} \times \mathrm{SEN}/(\mathrm{PRE} + \mathrm{SEN}), \tag{11}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

We first studied the best weighting scheme for feature concatenation in our proposed fused-feature-based CNN and its outperformance compared with the single model feature based CNNs. Then we compared the performances of model-dependent GMM-HMM method with data-dependent fused-feature-based CNN method.

A. Performance of 1D/2D-fused-feature-based CNN

As 1-D feature and 2-D feature are concatenated in our proposed network, we studied the importance of roles they play in AS detection, by attempting various weights. Fig. 6 shows the experimental results with different weights. It is shown that the best performance was achieved when 1-D and 2-D features were assigned 0.5 weights equally, which will be retained in the experiments that follow.

To show why a 1D/2D-fused-feature-based CNN would be preferred rather than 1-D or 2-D feature only CNN in implementing the addressed task, we also trained and tested these two CNNs. The indicators of performances of our proposed method as well as the compared methods are listed in Table 1. It is shown that, because many samples labeled as AS were judged as normal, the ACC and SEN under the 1-D-feature-based network failed to reach 90%. Compared with the 1-D-feature-based network, the 2-D-feature-based network yielded even worse performance. By contrast, all the performance indicators of our proposed method were larger than 90%, and PRE even reached 95%. This is not surprising, as the proposed 1D/2D-fused-feature-based CNN not only retains important information in the signal waveform in 1-D feature, but also grasps the auditory characteristics adapted to human ear by MFCC-based 2-D features.

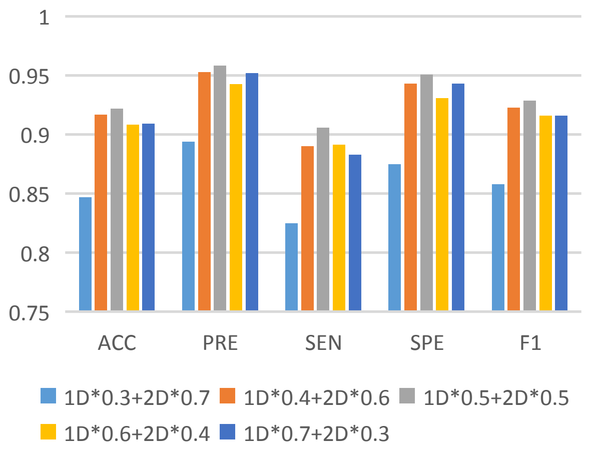

Fig. 6. Performances of our proposed neural network with different weights.

TABLE I. Performance Comparison of CNNs

Methods	Indicators of Performance				
	ACC	*PRE*	*SEN*	*SPE*	*F1-score*
1-D only	89.56%	92.98%	88.03%	91.51%	90.43%
2-D only	84.67%	89.38%	82.46%	87.51%	85.78%
Proposed	92.21%	95.29%	90.57%	94.29%	92.87%

B. Comparison between GMM-HMM and proposed CNN

In this paper, we proposed two machine learning methods for AS detection. It is interesting to learn the performance comparison between the model-dependent method and data-dependent method. The performance indicators are listed in Table II, where in this study the number of Gaussian distributions is set to be 3. It can be observed that the proposed 1D/2D-fused-feature-based CNN outperformed GMM-HMM-based method substantially, implying that when we have samples in plenty for a PCG classification task, a data-dependent method may be more preferable.

TABLE II. Performances of Two Proposed Methods

Methods	Indicators of Performance				
	ACC	*PRE*	*SEN*	*SPE*	*F1-score*
GMM-HMM	78.78%	74.59%	78.39%	79.08%	77.49%
1D/2D fused CNN	92.21%	95.29%	90.57%	94.29%	92.87%

In this paper, experiments were also carried out to study the effects of different size of training data set on the performances of GMM-HMM and 1D/2D-fused-feature-based CNN. 5000, 4000, 3000, and 2000 cardiac sound segments were randomly selected from the original training data set as the new training data set, and the original test set is always retained as the test set. The experimental results are shown in Fig. 7. It can be seen that, as the amount of samples in the training set decreases, the various indicators of GMM-HMM fluctuated within a small range, while these indicators of CNN showed a slow downward trend, and the performances of two methods were gradually approaching. But overall, the indicators of CNN are still higher than those of GMM-HMM even at 2000 samples in the training set, which further demonstrated the advantage achieved by the proposed 1D/2D-fused-feature-based CNN. In addition, from the perspective of the amount of parameters and training speed in model establishing, in GMM-HMM only the state transition probability matrix and observation probability matrix are studied, while a large number of parameters in the network need to be trained in CNN. Hence, in real applications, GMM-HMM may be easier to be developed in an electronic stethoscope than CNN.

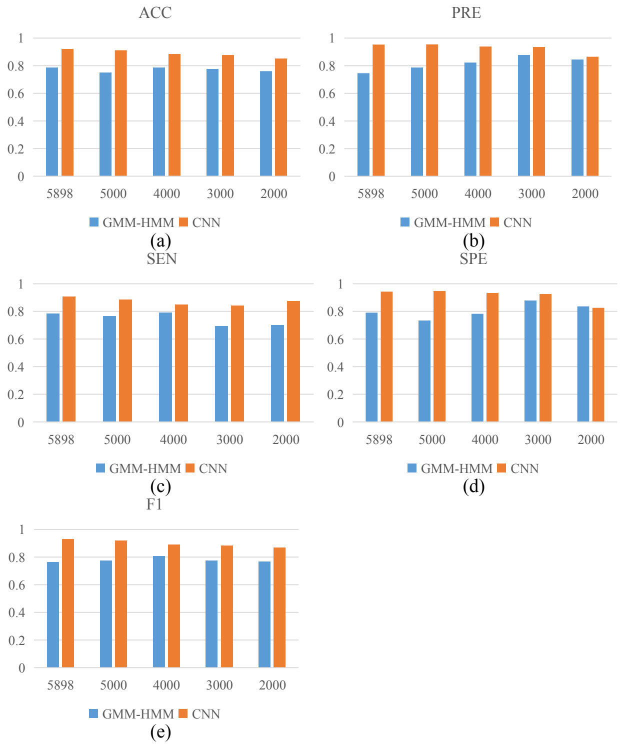

Fig. 7. Experimental results under different amounts of data in the training set.

VI. Conclusion

In this paper, machine learning methods for AS detection in electronic stethoscope were studied. GMM-HMM and 1D/2D-fused-feature-based CNN, corresponding to model-dependent and data-dependent methods, respectively, were proposed. The performances of the two proposed methods in AS detection were evaluated via real recorded PCG data collection, and their feasible scenarios were illustrated. The limitation of this study is that the AS detection is applied in each cardiac sound segment. Further study will combine AS detection with automatic cardiac sound segmentation.

References

[1] M. R. Montinari, and S. Minelli, "The first 200 years of cardiac auscultation and future perspectives," *J. Multidiscip. Healthc.*, vol. 12, pp. 183–189, Mar. 2019.

[2] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart Sounds–a systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.

[3] S. Choi, and Z. Jiang, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Syst. Appl.*, vol. 35, pp. 1679-1687, Nov. 2008.

[4] B. Al-Naami, J. Al-Nabulsi, and H. Amasha, "Utilizing wavelet transform and support vector machine for detection of the paradoxical splitting in the second heart sound," *Med. Biol. Eng. Comput.*, vol. 48, pp. 177-184, 2010.

[5] W. Zhang, X. Guo, and Z. Yuan, "Heart sound classification and recognition based on EEMD and correlation dimension," *J. Mech. Med. Biol.*, vol. 14, pp. 777-244, 2014.

[6] L. D. Avendaño-Valencia, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Feature extraction from parametric time-frequency representations for heart murmur detection," *Ann. Biomed. Eng.*, vol. 38, pp. 2716-2732, 2010.

[7] T. R. Reed, N. E. Reed, and P. Fritzson, "Heart sound analysis for symptom detection and computer-aided diagnosis," *Simul. Model. Pract. Th.*, vol. 12, pp. 129-146, May. 2004.

[8] K. Higuchi, K. Sato, H. Makuuchi, A. Furuse, S. Takamoto, and H. Takeda, "Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique," *J. Med. Eng. Technol.*, vol. 30, pp. 61-68, 2006.

[9] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, "Heart sound classification using deep structured features," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 565-568.

[10] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 621-624.

[11] V. Maknickas, and A. Maknickas, "Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiol. Meas.*, vol. 38, pp. 1671-1684, 2017.

[12] T. Chen, S. Yang, and L. Ho, "S1 and S2 heart sound recognition using deep neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, pp. 372-380, 2016.

[13] M. Rouhani, and R. Abdoli, "A comparison of different feature extraction methods for diagnosis of valvular heart diseases using PCG signals," *J. Med. Eng. Technol.*, vol. 36, pp. 42–49, 2012.

[14] C. Thomae, and A. Dominik, "Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 625-628.