

Follow the Sound of Children's Heart: A Deep-Learning-Based Computer-Aided Pediatric CHDs Diagnosis System

Bin Xiao^{ID}, Yunqiu Xu, Xiuli Bi^{ID}, Weisheng Li^{ID}, Zhuo Ma^{ID}, Junhui Zhang, and Xu Ma

Abstract—Auscultation of heart sounds is a noninvasive and less costly way for congenital heart disease (CHD) diagnosis, especially for pediatric individuals. The deep-learning-based computer-aided heart sound analysis has been widely studied and developed in recent years. In this article, we develop a deep-learning-based computer-aided system for pediatric CHDs diagnosis using two novel lightweight convolution neural networks (CNNs). One key issue of most existing deep-learning-based systems is the scarcity of large-scale data sets for CNN learning. To this end, we collect heart sounds from newborns and children with physicians' annotations to construct a pediatric heart sound data set that contains 528 high-quality recordings (nearly 4 h in total) from 137 subjects. With the constructed data set, deep CNN models can be easily trained as classifiers in computer-aided CHDs diagnosis systems. The experimental results demonstrate the superiority of our proposed methods in terms of diagnosis performance and parameter consumption in the application of Internet of Things.

Index Terms—Computer-aided diagnosis, convolution neural networks (CNNs), heart sound, pediatric congenital heart diseases (CHDs).

I. INTRODUCTION

CARDIOVASCULAR diseases (CVDs) continue to be the major cause of morbidity and mortality worldwide. An estimated 17.7 million people died from CVDs in 2015, representing 31% of all global deaths [1]. Timely recognizing and properly curing pediatric congenital heart diseases (CHDs) in the early stage are considered as one of the best ways to

Manuscript received November 2, 2019; revised December 10, 2019; accepted December 17, 2019. Date of publication December 20, 2019; date of current version March 12, 2020. This work was supported in part by the National Key Research and Development Project under Grant 2016YFC1000307-3, in part by the National Natural Science Foundation of China under Grant 61976031, in part by the Chongqing Natural Science Foundation under Grant cstc2018jcyjAX0117, and in part by the Scientific Technological Key Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201800601. (Corresponding authors: Zhuo Ma; Junhui Zhang.)

Bin Xiao, Yunqiu Xu, Xiuli Bi, and Weisheng Li are with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xiaobin@cqupt.edu.cn; imyunqiuwu@gmail.com; bixl@cqupt.edu.cn; liws@cqupt.edu.cn).

Zhuo Ma is with the School of Cybernetics, Xidian University, Xi'an 710126, China (e-mail: mazhuo@mail.xidian.edu.cn).

Junhui Zhang is with the Health Management Center, the First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China (e-mail: 2275610878@qq.com).

Xu Ma is with the Human Genetics Resource Center, National Health Commission, Beijing 100181, China.

Digital Object Identifier 10.1109/JIOT.2019.2961132

alleviate this situation, which not only makes individuals suffer less but also reduces medical costs. The cardiovascular system can be efficiently evaluated using some heart-related signals [55]: heart sound signals, heart electronic signals [22], mechanical vibrations signals [12], etc. Among these signals, auscultation of heart sounds or phonocardiograms (PCGs) is a less costly, noninvasive and effective way for preliminary CHDs detection, which reveals many pathologic conditions, such as arrhythmias, valve disease, heart failure, etc., [25]. However, the insufficiency of cardiologists remains a challenge, especially in low- and middle-income areas. Besides, physicians usually require a long-time training to obtain accomplished clinical skills and extensive subjective experiences for accuracy auscultation diagnosis. Therefore, developing an objective and efficient computer-aided tool for automatic heart sound analysis has become a promising research field.

Computer-aided heart sound analysis is not a new topic. It was first introduced by Gerborg *et al.* [15] who automatically classified the pathology of pathologic heart sounds from children with rheumatic heart diseases. A computer-aided heart sound analysis system can be usually considered as a threefold task: 1) segmentation; 2) feature extraction; and 3) classification. In the segmentation, each heart sound recording can be segmented into a series of fundamental heart sounds (FHSs) or further precisely segmented into state sequence [46], i.e., successive the first (S1), systolic, the second (S2) and diastolic heart sounds. A higher accuracy rate of segmentation usually encourages better performance in both feature extraction and classification. In the second step, various feature extraction algorithms [7], [16], [39], [54] can be leveraged to extract feature representation of heart sounds for subsequent classification. In the last step, based on the extracted features, various types of classifiers are trained according to the different types of extracted features. Generally speaking, conventional handcrafted feature-based approaches are fast to train. However, they usually depend on some complex pre- and post-processing procedures. Recently, deep-learning-based methods, especially convolutional neural networks (CNNs), with strong capabilities and functionalities, have achieved tremendous success in many practical tasks [33], [44]. There are also some existing works investigating deep-learning-based heart sound analysis [14], [30], [34], [37], [38], [50]. In the heart sound classification task, deep-learning-based methods unify feature extraction and subsequent classification steps into a combined end-to-end framework. This data-driven approach

can extract feature automatically via a learning way without complex hand-craft feature design procedures. However, deep-learning-based methods usually meet a data hungry issue. A large scale of annotated data is required for training deep-learning-based models.

Auscultation probably is the most appropriate approach to diagnose cardiac diseases from newborns or children, because of its noninvasive, convenience and less costly. Nonetheless, developing an automatic pediatric CHDs diagnosis system via heart sound analysis remains a challenging task. Most existing works [21], [35], [43] used handcrafted feature-based pipelines, evaluated on several small data sets with the limited number of recordings and very short heart sound durations, while deep-learning-based methods are still very difficult to be trained well, based on most existing data sets. Moreover, parameter-saving, yet effective, deep learning models for heart sound analysis are desired, especially in the application of Internet of Things (IoT).

Guided by this intuition, we collect a mass of heart sound recordings from newborns and children in the real clinical environment. At the same time, corresponding annotations are also provided to organize a pediatric heart sound data set for the first time. Based on this data set, we develop a deep-learning-based computer-aided pediatric CHDs diagnosis system using two different CNN models. Both of these CNN models obtain state-of-the-art results using extremely low parameter consumption, which can be used in the application of IoT.

To sum up, our main contributions are threefold.

- 1) We first collect and organize a new publicly available large-scale pediatric heart sound data set, so that deep-learning models can be trained for pediatric CHDs diagnosis.
- 2) We develop a deep-learning-based computer-aided diagnosis system for pediatric CHDs diagnosis, which focuses on improving the feature reuse ability and lowering the parameter consumption of CNN models.
- 3) Experiments demonstrate the superiorities of our proposed system comparing with other state-of-the-art CNN-based methods in terms of diagnosis performance and parameter consumption.

The remainder of this article is structured as follows. In Section II, we discuss some previous works related to ours. The new constructed pediatric heart sound data set is described in Section III. The implementations of proposed deep-learning-based computer-aided CHDs diagnosis system are introduced in Section IV. In Section V, we empirically evaluate the developed system on our pediatric heart sound data set, showing superiorities of our proposed methods in terms of diagnostic accuracy and the cost of parameters. We also investigate the interpretability of our trained CNN model through additional experiments. Section VI concludes this article.

II. RELATED WORK

Heart sound analysis has been widely studied for a long time. Segmentation is the first step in conventional automatic heart sound analysis, which can mainly be grouped into envelope-based [48], feature-based [32],

machine learning-based [49], hidden Markov model (HMM)-based [46], and deep-learning-based methods [10], [28], [29]. In the majority of heart sound analysis systems, feature extraction is the most important and crucial step [13]. After segmentation, the time [7], frequency [39] or time-frequency features [16], [54] are extracted to reveal some physiological and pathological information for subsequent classification. In the last step, the extracted discriminative features are utilized to train the machine learning-based classifiers, such as artificial neural network (ANN) [51], support vector machines (SVMs) [27], HMM [41], k-nearest neighbor (k-NN) [36], etc. Automatic analysis of heart sounds from children or newborns is a more challenging task, there were few works which studied the pediatric heart sound analysis in recent years. Pretorius *et al.* [35] segmented pediatric heart sounds with the help of synchronous ECGs. Then, they extracted features and classification using principal component analysis (PCA) and an ensemble of ANNs, respectively. Kang *et al.* [21] automatically identified the innocent still's murmur from all other murmurs in children's heart sounds using SVM classifier based on 14 types of temporal and spectral features. Based on the Arash-Band feature extraction method and SVM, Sepehri *et al.* [43] developed an automatic pediatric cardiac disease screening method on Android-based devices.

Except for the conventional machine learning-based methods, there are also some attempts using deep-learning-based methods that are able to merge feature extraction and classification into an end-to-end model and extract the most discriminative features automatically. CNNs as the typical deep-learning-based method have been used in automatic heart sound classification [6], [14], [30], [34], [37], [38], [50], [52]. Ryu *et al.* [38] fed 1-D heart sound patch into a simple 1-D CNN after using Windowed-sinc Hamming filter. In [14], [30], and [37], the heart sound classification task were treated as image classification, and they extracted 2-D maps based on Mel-frequency cepstral coefficients (MFCCs), power spectral density (PSD), and neuromorphic auditory sensor, respectively. Furthermore, CNN and recurrent neural network (RNN) hybrid models were also introduced. Thomae and Dominik [50] adopted an RNN with a CNN front-end for heart sound classification, while most recently Alam *et al.* [6] introduced a parallel recurrent and CNN for heart sound murmur detection. In 2016, Potes *et al.* [34] achieved the first prize of PhysioNet/Computing in Cardiology 2016 Challenge (PhysioNet/CinC 2016) [25], where they used a combination of a decomposed-band CNN and ensemble of a variant AdaBoost. However, to the best of our knowledge, there are no prior works focusing on pediatric heart sound diagnosis based on deep learning methods, because of the absence of large-scale pediatric heart sound data set.

As a kind of data-driven approach, deep-learning-based models are usually data hungry and require a mass of well-annotated data. PhysioNet/CinC 2016 first provided a large-scale heart sound data set with 3126 annotated heart sound recordings, and each recording lasts from 5 s to just over 120 s. The sufficient recordings provided opportunities to train a deep-learning model for heart sound analysis. According to Liu *et al.* [25], there were no previous works

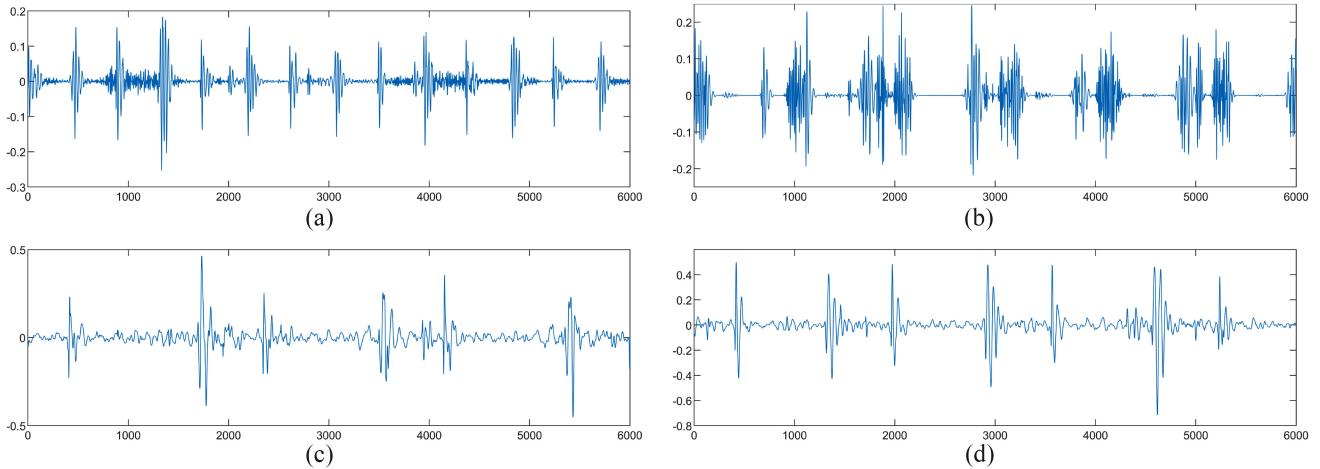


Fig. 1. Heart sound samples collected from different subjects (3 s). (a) Normal pediatric heart sound from our data set. (b) VSD pediatric heart sound from our data set. (c) Normal adult heart sound from [25]. (d) Abnormal adult heart sound from [25].

using deep-learning-based methods for heart sound classification before PhysioNet/CinC 2016, since the lack of training samples with high quality and quantity. Before [25], only a small number of heart sound data sets were publicly available: the Michigan heart sound and murmur library (MHSML) [2], the PASCAL Classifying Heart Sounds Challenge 2011 data set (PASCAL 2011) [8], Washington heart sounds and murmurs data set (WHSM) [4], and the Cardiac auscultation of heart murmurs data set (CAHM) [3] that requires payment for access. Among these data sets, they all have some limitations in terms of numbers or length of recordings, which are insufficient to train a deep-learning-based model well. It also can be observed that among all the existing data sets, heart sound recordings were only collected from adults or adults and children. This is because the heart sound collection from newborns and children has its difficulties.

III. PEDIATRIC HEART SOUND DATA SET

As discussed in the previous sections, the lack of training data impedes the deep-learning-based methods to computer-aided heart sound analysis tasks, especially for the pediatric heart sound analysis. Since the processes of both pediatric heart sound collection and annotation are quite costly and labor-intensive, to the best of our knowledge, there is not any publicly available data set to deal with this issue so far.

Collecting heart sounds from adults is much easier than from newborns and children. Adult subjects can make less unrelated noise by following the physicians' instructions, while pediatric subjects usually bring extra noise. For instance, the noise generated by children's crying, yelling, cough, intestinal peristalsis, and moving are usually unavoidable. Additionally, the heart sounds collected from adult subjects cannot be directly used to train a model for pediatric heart sound analysis, because of the nonpathological variations between children's and adults' heart sounds, including faster heart rate, physiological third heart sounds (S3), etc. Fig. 1(a) and (b) shows the heart sounds collected from a healthy pediatric subject and a child with ventricular septal defect (VSD). For comparison, a normal and an abnormal recording from PhysioNet/CinC 2016 are also shown in the bottom of Fig. 1.

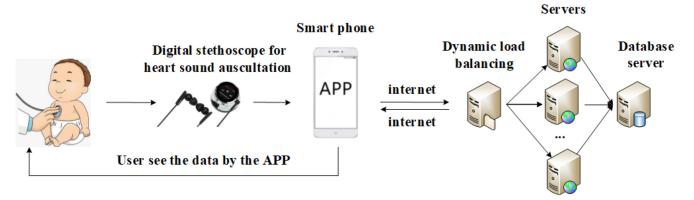


Fig. 2. Flow diagram of our computer-aided diagnosis system.

Inspired by the heart rate monitoring system [24], we developed a pediatric CHDs diagnosis system and distributed it in the cooperation children hospital. The main pipeline of our proposed system is shown in Fig. 2, which consists of a digital stethoscope, an IoT device, and cloud servers. The Web, Android or iOS APPs could be equipped in the IoT device, such as a smart phone. The server program language is Java and the database is MySQL. The diagnosis process can be performed on both the cloud servers and the users' smart phone depending on the Internet resource. If the Internet resource is limited, the diagnosis is only performed on the users' smart phone. On the contrary, if the Internet resource is abundant, users' information can be upload to the cloud servers to have the diagnosis and obtain the results feedback on their smart phone. This system is supported by our constructed publicly available data set¹ with pediatric heart sounds as well as their corresponding annotations and age information.

In total, 528 heart sound recordings are collected from 137 subjects and well annotated by professional physicians in our cooperating hospital. All the heart sounds in our data set are collected from pediatric subjects aged from 1 months to 12 years old. All of the recordings are collected in the real clinical environment using the Thinklabs One digital stethoscopes [5] with a sampling frequency of 44.1 kHz and 16 bits per sample. All the heart sounds are grouped into seven categories: normal, atrial septal defect (ASD), VSD, both ASD and VSD, tetralogy of Fallot (TOF), both ASD and TOF, and other heart-related diseases, such as mitral regurgitation, aortic

¹This pediatric heart sound data set collection has passed the hospital ethics committees.

TABLE I
SUMMARY OF EXISTING PUBLICLY AVAILABLE HEART SOUND DATA SETS

| Heart sound dataset | # recordings | Data resources | Annotation categories |
|--------------------------|--------------|-----------------------|--|
| MHSML [2] | 23 | Adults | Normal, mitral valve prolapse, innocent murmur, etc. |
| PASCAL 2011 [8] | 656 | Unknown | Normal, murmur, extrasystole, artifact, etc. |
| CAHM [3] | 64 | Unknown | Normal, ASD, VSD, diastolic aortic regurgitation, etc. |
| WHSN [4] | 16 | Unknown | Normal, Aortic stenosis, ASD, VSD, etc. |
| PhysioNet/CinC 2016 [25] | 3126 | Adults and children | Normal, abnormal |
| Ours | 528 | Newborns and children | Normal, ASD, VSD, ASD & VSD, TOF, etc. |

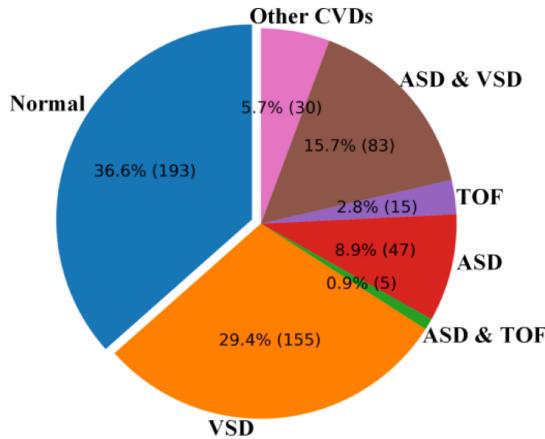


Fig. 3. Data distribution of our constructed pediatric heart sound data set.

stenosis, pulmonary stenosis, etc. The data distribution of our pediatric heart sound data set is depicted in Fig. 3. For data cleaning purpose, we also manually cut out some heart sound patches with too many distinctly unrelated noises, such as voices from children's parents, crying and yelling from children, etc. In summary, 528 pediatric heart sound recordings (nearly 4 h in total) are kept in our data set, and each recording is 3–249 s in duration. Table I shows the comparison of our data set and other existing publicly available heart sound data sets.

The performance of deep-learning-based methods usually relies on continuous data collection. Collecting and leveraging the data often brings a number of privacy issues, especially in the medical related domains. In order to preserve individual's privacy, the diagnosis can be completely performed on users' smart phone. The diagnosis also could be done on the cloud server, while their information is uploaded with cryptographic techniques [9], [45]. Additionally, users could use their IoT devices to run initial layers of networks to generate intermediate results, before feeding individual data to the cloud server [31].

IV. DEEP-LEARNING-BASED PEDIATRIC HEART SOUND CLASSIFICATION

With the constructed data set, we develop the 1-D CNNs-based computer-aided pediatric CHDs diagnosis system using heart sound in this section. The overall pipeline of our designed system is depicted in Fig. 4. We first describe the heart sound preprocessing and segmentation. Then the details

of heart sound patches classification using 1-D CNNs are provided. At last, the final decision rule transforming the results of patch-level into recording-level diagnosis is introduced.

A. Preprocessing and Segmentation

Transferring 1-D raw signals to 2-D representations such as MFCC is the most common choice in many state-of-the-art CNN-based methods. However, a set of hyperparameters are required to optimize the transformation, although the 2-D representations sometimes represent acoustical patterns well. In addition, much more parameters will be cost when CNN models are trained with 2-D convolution kernels. Therefore, the 1-D raw waveform data are directly used as the input in our system. The raw heart sounds are resampled to 2000 Hz to eliminate some unrelated noises and downsize the inputs for subsequent training.

As mentioned above, good segmentation of heart sounds can boost the performance of subsequent feature extraction and classification in some cases. However, accurate segmentation remains a challenging task and brings extra computing. Based on the assumptions (the abnormal heart sound recording can be determined in few seconds; and the abnormality of heart sounds can be observed in every beat if there is a presence of heart pathology) introduced in [30], we adopt the sliding window with a fixed length to segment each heart sound recording into several patches, which is also more suitable for our CNN-based structures. A window length of three seconds is empirically selected, and every patch is labeled according to the category of the corresponding recording.

B. Heart Sound Patch Classification

In this section, the heart sound patch classification within our diagnosis system is introduced. More specifically, two novel low parameter-consumed 1-D CNN architectures are proposed for heart sound classification. The overview of two networks is demonstrated in Fig. 5. The main variation between the two models is that two different basic blocks are stacked, i.e., the dense block-based model and the clique block-based model.

Dense Block: The densely connected convolutional network [20] with its advantage in feature reuse has achieved a great number of promising results in image recognition. Based on this, a novel 1-D dense block structure is developed by us for pediatric heart sound patches classification. As illustrated in left column of Fig. 6, in our first architecture, layers are all densely connected in each dense block. In a dense block, subsequent layers take the combination of all their preceding

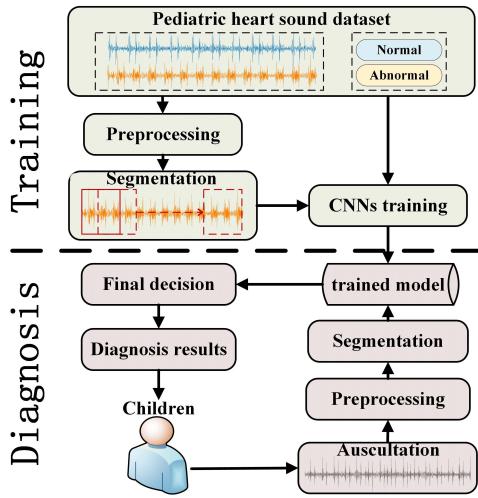


Fig. 4. Flow diagram of our computer-aided diagnosis system.

layers produced feature maps as inputs. Formally, we define the $H(\cdot)$ as a composite function of a series of operations, such as convolutional layers, pooling layers, batch normalization layers, or activate functions. In a conventional network, feature maps produced by the l th layer can be formulated as (1), while these in a dense block are demonstrated in (2), where x_0 indicates the input feature maps of each block and $[\cdot]$ denotes the concatenation operation of feature maps. As a result, the densely connected features can be viewed as a global state of this block, which enables layers in every block not only take all their preceding layers to produce features for their subsequent layers, but also back propagate gradients to their previous layers directly

$$x_l = H_l(x_{l-1}) \quad (1)$$

$$H([x_0, x_1, \dots, x_{l-1}]) \quad (2)$$

$$x_l^{(i)} = \begin{cases} H([x_0, x_1^{(i)}, \dots, x_{l-1}^{(i)}]), & i = 1 \\ H([x_1^{(i)}, \dots, x_{l-1}^{(i)}, x_{l+1}^{(i-1)}, \dots, x_L^{(i-1)}]), & i = 2. \end{cases} \quad (3)$$

Clique Block: In our second architecture, clique blocks [53] that also focus on feature reuse are adopted as basic blocks to build our network, rather than dense blocks. On the basis of dense block, an extra propagation step is developed in every block, constructing a bidirectional connection to further enhance the feature reuse. In the first step of the bidirectional connection, the propagation in each clique block is identical with that in a dense block. Unlike dense block, another feature map update propagation step is designed in the clique block, shown in the middle column of Fig. 6. For each layer in clique block, not only the preceding feature maps in the second step are used but the subsequent feature maps in the first step are taken into consideration. Formally, this two-step propagation is formulated as (3), where l indexes the layer of a clique block with L -layer and i indicates different step of propagation.

Separable Convolution: The aim of ours is to develop a lightweight computer-aided heart sound diagnosis system. In order to further lower the parameter consumption, the

separable convolutions [11], [18], [40] are also utilized. The 1-D separable convolutions can be separated into two steps, i.e., a depth-wise convolution followed by a point-wise convolution, which can be formulated as follows:

$$\text{PointConv}(W, y)_{(i)} = \sum_{c=1}^C W_c \cdot y_{(i,c)} \quad (4)$$

$$\text{DepthConv}(W, y)_{(i)} = \sum_{k=-K}^K W_k \odot y_{(i+k)} \quad (5)$$

$$\text{SepConv}(W_p, W_d, y)_{(i)} = \text{PointConv}_{(i)} \times (W_p, \text{DepthConv}_{(i)}(W_d, y)) \quad (6)$$

where \odot denotes the element-wise product, $W \in \mathbb{R}^{(2K+1) \times C}$ denotes the weights of the convolutional kernels, and y denotes the input feature maps. The separable convolutions that can significantly reduce the cost of parameters are used in every composite layer instead of the conventional convolutions. Different to previous architectures [11], [18], [40] with separable convolutions, dense connections are introduced rather than residual connections. The dense connection can back propagate the information flow more efficiently with fewer parameters.

Transition Block: When the number of layers increases, the number of direct connections grows quadratically, which is very memory demanding. In this article, the transition blocks with pooling layers are introduced to divide the networks into several blocks processing feature maps at different resolutions. Furthermore, a channel-wise attention mechanism [19] is also suggested in every transition block. As shown in Fig. 7, the mechanism can be separated into two steps. In order to exploit wider contextual information outside of the receptive field in lower layers, in the first squeeze step, the spatial information is squeezed into a channel descriptive vector using a global average pooling (GAP) layer. In the second excitation step, the channel-wise dependencies are obtained based on a simple gating mechanism with a sigmoid activation. Then, the excited descriptive vector that measures the channel-wise dependencies is multiplied by the input feature maps at corresponding channels. This mechanism performs a dynamic feature recalibration, which increases the sensitivity of the network to informative features and weaken noises for ensuring that more useful features can be exploited efficiently by the subsequent layers.

C. Final Decision Rule

The ultimate objective of our computer-aided system is to diagnosis pediatric heart sounds into different categories based on the given heart sound recordings. To this aim, the final recording-level diagnosis results are supposed to be provided to users, based on the classification results of heart sound patches. As mentioned in Section IV-A, the abnormality of heart sounds can be observed in every beat if there is a presence of heart pathology. Therefore, the majority voting strategy is applied to give the final diagnosis results. In this majority voting, the number of labeled patches from each recording is counted. If the number of patches with any categories is

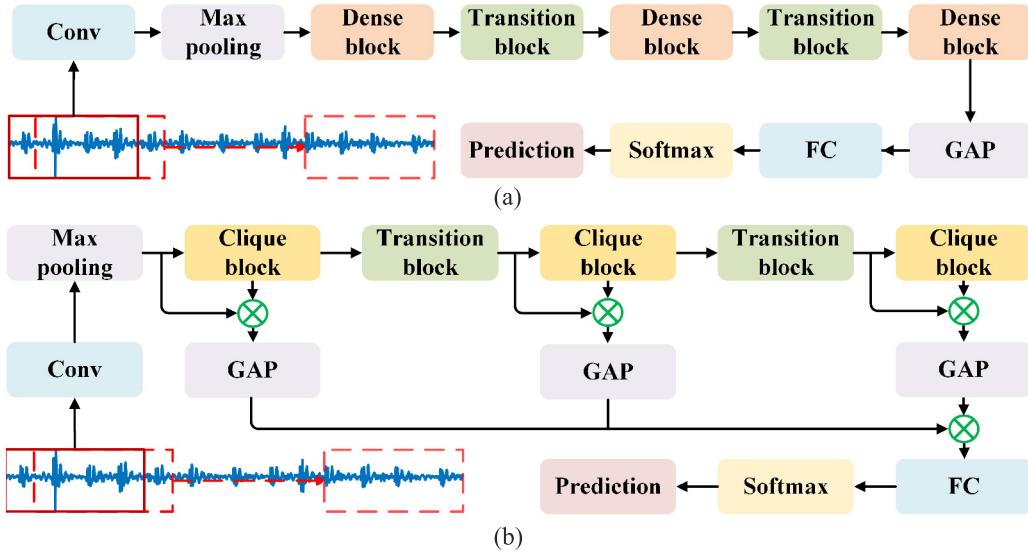


Fig. 5. Overview of two networks. \otimes in figures indicates the concatenation of feature maps. (a) Dense block-based architecture. (b) Clique block-based architecture.

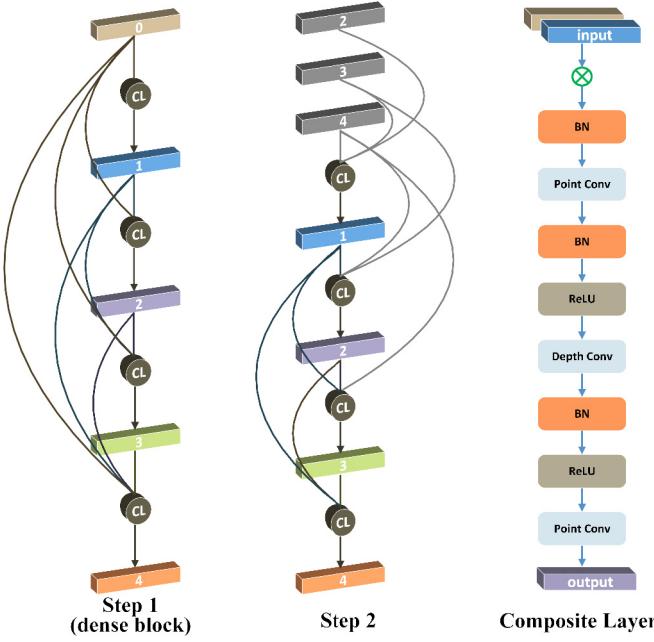


Fig. 6. Architecture of clique and dense block. The first-step propagations in clique blocks are identical to that in dense blocks. Feature maps colored in gray refer to feature maps generated in stage 1. The CL represents a composite layer illustrated on the right.

larger than others, the recording can be diagnosed as this category. When the maximum numbers of different categories for a single recording are equal, the mean of original predicted probability values of patches will be compared.

V. EXPERIMENTS

In this section, several experiments on our proposed pediatric heart sound data set are conducted to verify the effectiveness of our proposed CHDs diagnosis system. Due to the number of heart sound recordings in our data set for some

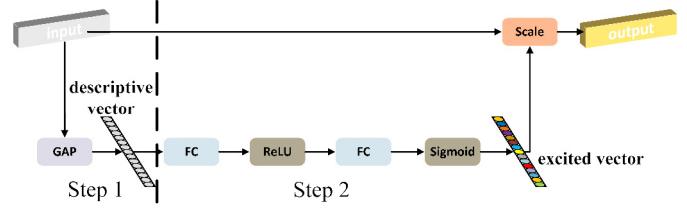


Fig. 7. Pipeline of the attention mechanism in transition blocks.

specific CHDs is rare, it is still not sufficient to train a deep-learning-based model for multiclass diagnosis. So, we further group the collected pediatric heart recordings into two categories, establishing a normal and abnormal binary diagnosis task. In order to show the generalization of our proposed system, experiments on another large-scale heart sound data set PhysioNet/CinC 2016 [25] are also performed in this section.

A. Implementation Details

Based on the constructed pediatric heart sound data set, we empirically segment the 1-D waveform heart sounds recordings into three-second length patches with a stride of 1 s as the input of our computer-aided pediatric heart sound diagnosis system.

As depicted in Fig. 5, in both block-based networks, each architecture is mainly stacked by three basic blocks and transition blocks between them. A convolution layer with a relatively larger kernel size is first applied to extract more abundant low-level features from input signals, rather than feeding inputs into the first block directly. After the convolution layer, a max pooling is also utilized to downsize the resolution of feature maps.

In the proposed dense block-based network, dense blocks with six composite layers and a growth rate of 12 are adopted. Each composite layer is consisting of a series operations as

TABLE II
ARCHITECTURE DETAILS OF DENSE BLOCK-BASED CNN. BN AND RELU LAYERS ARE NOT SUMMARIZED IN THIS TABLE

| Layers | Output size | Configurations |
|----------------------|------------------|--|
| Inputs | 6000×1 | - |
| Convolution layer | 3000×64 | conv 7, stride 2 |
| Pooling layer | 1499×24 | max pool 3, stride 2 |
| Dense block | 1499×96 | # CL 6, growth rate 12 separable conv 3, stride 1 |
| Transition block | 749×48 | point-wise conv attention mechanism average pool 2, stride 2 |
| Dense block | 749×120 | # CL 6, growth rate 12 separable conv 3, stride 1 |
| Transition block | 374×60 | point-wise conv attention mechanism average pool 2, stride 2 |
| Dense block | 374×132 | # CL 6, growth rate 12 separable conv 3, stride 1 |
| Pooling layer | 1×132 | GAP |
| Classification layer | 2 | FC, softmax |

illustrated in the right column of Fig. 6. In each transition block, an average pooling layer with a size of 2 is equipped after the attention mechanism. A GAP layer is followed by a fully connected layer and a Softmax. More details about this network are shown in Table II.

In the proposed clique block-based network, five composite layers with a growth rate of 12 are staking in each clique block. As depicted in Fig. 5(b), the overall architecture is also a little different to the dense block-based network, excluding the different type of basic blocks. Concretely, the feature maps produced by each clique block are not merely fed into subsequent block but also concatenated with the input feature maps of this block and compressed to half of the original channels. The compressed feature maps are concatenated and followed by a GAP layer. More architectural details about this network are shown in Table III.

For two networks, a cross-entropy loss with a weight of 0.25–1 (abnormal to normal) is empirically adopted as the loss function. SGD with 0.9 Nesterov momentum is utilized for optimization. Weight decay of 10^{-4} and dropout of 0.1 are set for regularization. The dense block-based and the clique block-based models are trained for 120 and 200 epochs with early stop. The initial learning rate of 0.1 and 0.01 are selected for dense block-based and clique block-based models, while the learning rate decay of 0.1 is implemented at epoch of 40 and 60 for dense block-based models, and at epoch of 100 and 150 for clique-block-based models.

B. Evaluation Metrics and Experimental Results

Several evaluation metrics are carried out to evaluate the performance of the proposed system: accuracy (Acc) as a statistical measure has been most commonly used in classification, which measures how well a classifier is; precision of normal (Pr_{normal}) and precision of abnormal (Pr_{abnormal}) are the proportions of positive results for two categories in a diagnosis, describing the performance of a diagnostic system; specificity (Sp) and sensitivity (Se) indicate the proportion of correctly identified negatives and positives; based on Sp and

TABLE III
ARCHITECTURE DETAILS OF CLIQUE BLOCK-BASED CNN. BN AND RELU LAYERS ARE NOT SUMMARIZED IN THIS TABLE

| Layers | Output size | Configurations |
|----------------------|------------------|--|
| Inputs | 6000×1 | - |
| Convolution layer | 3000×64 | conv 7, stride 2 |
| Pooling layer | 1499×64 | max pool 3, stride 2 |
| Clique block | 1499×60 | # CL 5, growth rate 12 separable conv 3, stride 1 |
| Pooling layer | 1×62 | GAP |
| Transition block | 749×60 | point-wise conv attention mechanism average pool 2, stride 2 |
| Clique block | 749×60 | # CL 5, growth rate 12 separable conv 3, stride 1 |
| Pooling layer | 1×60 | GAP |
| Transition block | 374×60 | point-wise conv attention mechanism average pool 2, stride 2 |
| Clique block | 374×60 | # CL 5, growth rate 12 separable conv 3, stride 1 |
| Pooling layer | 1×60 | GAP |
| Classification layer | 2 | FC, softmax |

Se, the average of these two values, i.e., the overall score (Score) is introduced to describe the comprehensive diagnosis performance. We consider the correctly diagnosed abnormal pediatric heart sound recordings as true positive samples in this article. Moreover, the consumption of trainable parameters (Params) of CNN is also calculated to evaluate the scale of each model.

C. Classification Results

In our constructed pediatric heart sound data set, our proposed methods are mainly compared with three state-of-the-art CNN-based heart sound classification methods [30], [37], [38]. The experimental results are summarized in Table IV.

It can be observed that the proposed dense block-based method outperforms other methods except Pr_{normal} and Se that are achieved by 1-D CNN [38]. Our proposed clique block-based method obtains the second place and the third place in terms of Pr_{abnormal} and Score using only 0.19 M parameters. However, the clique block-based method with 0.08 M more parameters than the dense block-based method does not excel in any aspect of metrics, which is different in some computer vision tasks. This is probably because the structure of the clique block-based method is overcomplex and the current scale of training data in our data set is insufficient to reveal its stronger ability.

It is also not surprising that PSD-CNN [30] obtains unsatisfactory results, because the smallest number of training data can be generated through their preprocessing approach, which may be insufficient to train their model well. Despite the largest number of training data and trainable parameters are used by MFCC-CNN [37], our proposed method still obtains the highest Score with a minimal Params that is less than one-hundredth of [37]. This is mainly on account of the strong ability in feature extraction and feature reuse of our proposed methods.

To further demonstrate the good generality of the proposed method, experimental results on PhysioNet/CinC 2016 [25]

TABLE IV
DIAGNOSIS RESULTS ON THE PEDIATRIC HEART SOUND DATA SET

| Method | Acc | Pr_{normal} | $Pr_{abnormal}$ | Sp | Se | Score | Params | # training patches |
|---------------|--------|---------------|-----------------|--------|--------|--------|---------|--------------------|
| 1-D CNN [38] | 0.9110 | 0.9258 | 0.9092 | 0.8106 | 0.9609 | 0.8858 | 0.19 M | 4860 |
| MFCC-CNN [37] | 0.9233 | 0.9114 | 0.9322 | 0.8825 | 0.9558 | 0.9192 | 12.41 M | 15836 |
| PSD-CNN [30] | 0.8566 | 0.8882 | 0.8422 | 0.6644 | 0.9597 | 0.8121 | 0.24 M | 2327 |
| Ours-dense | 0.9621 | 0.9167 | 0.9877 | 0.9808 | 0.9536 | 0.9672 | 0.11 M | 12045 |
| Ours-clique | 0.9168 | 0.8896 | 0.9366 | 0.8917 | 0.9340 | 0.9129 | 0.19 M | 12045 |

TABLE V
CLASSIFICATION RESULTS ON PHYSIONET/CINC 2016 DATA SET. RESULTS TAKEN FROM PREVIOUS WORKS ARE DENOTED BY \dagger

| Method | Acc | Pr_{normal} | $Pr_{abnormal}$ | Sp | Se | Score | Params | Precise segmentation |
|---------------------------------|--------|---------------|-----------------|--------|--------|--------|---------|----------------------|
| Feature extraction-based | | | | | | | | |
| SS-PLSR \dagger [43] | - | 0.9500 | 0.8200 | - | - | - | - | ✓ |
| SS-TD \dagger [21] | - | 0.9200 | 0.8800 | - | - | - | - | ✓ |
| CNN-based | | | | | | | | |
| 1-D CNN [38] | 0.8933 | 0.9364 | 0.7396 | 0.9282 | 0.9608 | 0.8445 | 0.19 M | ✗ |
| MFCC-CNN [37] | 0.9331 | 0.9538 | 0.8536 | 0.9516 | 0.8266 | 0.8891 | 12.41 M | ✓ |
| PSD-CNN [30] | 0.8905 | 0.9491 | 0.7103 | 0.9102 | 0.8150 | 0.8626 | 0.24 M | ✗ |
| AdaBoost-CNN \dagger [34] | - | - | - | 0.8200 | 0.8800 | 0.8500 | - | ✓ |
| DRGE \dagger [50] | - | - | - | 0.9600 | 0.8300 | 0.9000 | - | ✗ |
| Ours-dense | 0.9356 | 0.9609 | 0.8441 | 0.9573 | 0.8529 | 0.9051 | 0.11 M | ✗ |
| Ours-clique | 0.9328 | 0.9627 | 0.8290 | 0.9516 | 0.8621 | 0.9069 | 0.19 M | ✗ |

are reported in Table V. It can be found that the state-of-the-art results are also obtained by our proposed methods with fewer computational consumption, comparing with some advanced CNN-based methods [30], [34], [37], [38], [50] and handcrafted feature-based methods [21], [43]. Additionally, comparing with some of other methods utilizing the precise segmentation algorithm (such as [46]), a simple yet effective sliding window-based segmentation operation is only required in our pipeline.

D. Computational Complexity

In order to verify the efficiency of our proposed method, we design an experiment to test the runtime of our proposed method on a server with NVIDIA GTX TiTan PASCAL GPU and Intel Xeon E5-2603 CPU (1.7 GHz), and “Huawei Mate 20 Pro” smart phone with 8 GB memory and Hisilicon Kirin 980 CPU, respectively. The experiment is based on the Tensorflow framework, and the CNN models pretrained on the server are following the compiled on smart phone using Tensorflow Lite. We exclude the communication time between the users and server, and only record the runtime time of the CNN models. Moreover, for fair comparison, the experiment has been run ten times and the average elapsed time is recorded as the computational time of each method. The experimental results are listed in Table VI. It can be found that our proposed method takes little time for CNN inference on both the server and smart phone. Although a little bit of more parameters is required by the clique-based model, it uses less time for inference. We attribute this to more composite layers that are utilized in the dense-based model. Both two models can further be used in the application of IoT, such as some smart wearable devices or mobile usage. With more specialized optimization [17], [26] for CNN inference, such as TensorRT, the processing speed would be further boosted.

TABLE VI
COMPUTATIONAL COMPLEXITY COMPARISON ON SERVER AND SMART PHONE

| Method | seconds per patch | seconds per recording |
|--------------|-------------------|-----------------------|
| Server: | | |
| Ours-dense | 0.0206 | 0.4862 |
| Ours-clique | 0.0093 | 0.2198 |
| Smart phone: | | |
| Ours-dense | 0.4567 | 10.7879 |
| Ours-clique | 0.1502 | 3.5492 |

E. Visualization Results

Interpretation of the diagnosis results from the computer-aided diagnosis systems is a major issue that most physicians and patients concern about. Nonetheless, interpreting the results produced by deep learning models remains a challenging task. There are some works exploring the interpretability of deep-learning-based method in medical image [23] and ECG analysis [47]. But there is no prior work about heart sounds, because of the complexity and instability of heart sound signals.

In this visualization experiment, we attempt to explore the interpretability of our models through visualizing the class activation maps (CAMs) of each heart sound patch. Our implementations are mainly inspired by [42], and in order to obtain the cross-category visualization results, we normalized the gradients of both categories together. Only dense block-based network generated CAMs are illustrated in Fig. 8 as a concise example. The most and the least discriminative features among a heart sound patch are colored in red and blue, respectively, indicating what extent those areas influence our model toward a correct diagnosis. Generally speaking, for the correctly diagnosed category, higher active values can be observed from the whole heart sound patch. It is quite identical to the assumption that corresponding sounds can be found in every beat in general when the heart pathology exists [30]. It also can be noted that the largest proportion of the active areas for abnormal

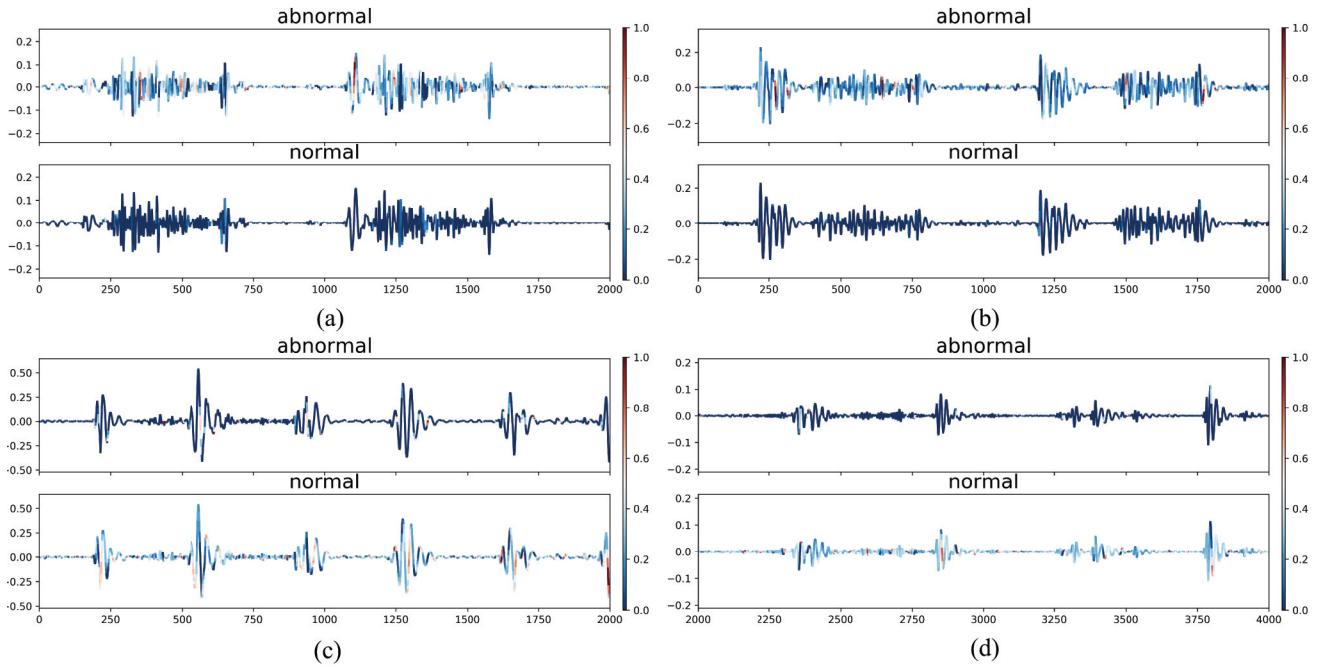


Fig. 8. Visualization results on our pediatric heart sound data set. Heart sounds collected from pediatric subjects with CHDs are depicted in (a) and (b), while healthy children's heart sounds are shown in (c) and (d).

heart sounds are more likely to be located in some discrete regions, such as S1, S2, or murmur heart sounds, indicating more discriminative features that our model mainly depends on might be contained in these areas.

VI. CONCLUSION

In this article, we first constructed a new pediatric heart sound data set which has 528 recordings collected from 137 subjects. All these recordings were collected and annotated by professional physicians in the real clinical environment. The availability of abundant recordings with their annotations allows developing a deep-learning-based computer-aided system for pediatric CHDs diagnosis. On this basis, we developed a lightweight pediatric CHDs diagnosis system using two different 1-D CNNs-based models that are all focus on feature reuse and parameter efficiency. The experimental results show that the developed system with significantly low parameter consumption outperforms state-of-the-art CNN-based methods, and can be used as a monitoring system for preliminary CHDs detection.

The smart device for healthcare is a promising application. In our future works, models with less diagnosis time consumption and higher diagnostic accuracy will be further investigated. At the same time, our built pediatric heart sound data set is supposed to be enlarged in terms of recordings' quantity and diversity of subjects. Based on this data set, diagnosis of different CHDs will also be studied.

REFERENCES

- [1] (May 17, 2017). *WHO, Cardiovascular Diseases (CVDs)*. Accessed Oct. 3, 2017. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] R. D. Judge. (Apr. 14, 2015). *Heart Sound & Murmur Library*. Accessed: Feb. 17, 2018. [Online]. Available: <https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library>
- [3] *Cardiac Auscultation of Heart Murmurs*. Accessed: Feb. 17, 2018. [Online]. Available: <http://www.egeneralmedical.com/listohearurm.html>
- [4] L. E. Pinsky and J. E. Wipf. *Heart Sounds & Murmurs*. Accessed: Feb. 16, 2018. [Online]. Available: <https://depts.washington.edu/physdx/heart/demo.html>
- [5] *Thinklabs One Digital Stethoscope*. Accessed: Dec. 9, 2017. [Online]. Available: <https://www.thinklabs.com/>
- [6] S. Alam, R. Banerjee, and S. Bandyopadhyay, "Murmur detection using parallel recurrent & convolutional neural networks," in *Proc. ACM KDD workshops*, Aug. 2018.
- [7] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, 2010.
- [8] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. (2011). *The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results*. [Online]. Available: <http://www.peterjbentley.com/hearthallenge/index.html>
- [9] D. Chen and H. Zhao, "Data security and privacy protection issues in cloud computing," in *Proc. IEEE Int. Conf. Comput. Sci. Electron. Eng.*, vol. 1, 2012, pp. 647–651.
- [10] T. Chen *et al.*, "S₁ and S₂ heart sound recognition using deep neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, Feb. 2017.
- [11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
- [12] T. Choudhary, L. N. Sharma, and M. K. Bhuyan, "Heart sound extraction from sternal seismocardiographic signal," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 482–486, Apr. 2018.
- [13] G. D. Clifford *et al.*, "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, no. 8, p. E10, 2017.
- [14] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, "Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 1, pp. 24–34, Feb. 2018.
- [15] D. S. Gerbarg, A. Taranta, M. Spagnuolo, and J. J. Hofler, "Computer analysis of phonocardiograms," *Progr. Cardiovascular Diseases*, vol. 5, no. 4, pp. 393–405, 1963.

- [16] M. Hamidi, H. Ghassemian, and M. Imani, "Classification of heart sound signal using curve fitting and fractal dimension," *Biomed. Signal Process. Control*, vol. 39, pp. 351–359, Jan. 2018.
- [17] J. Hanhirova, T. Kämäriinen, S. Seppälä, M. Siekkinen, V. Hirvisalo, and A. Ylä-Jääski, "Latency and throughput characterization of convolutional neural networks for mobile computer vision," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 204–215.
- [18] A. G. Howard *et al.*, "Mobilennets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.
- [20] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [21] S. Kang, R. Doroshow, J. McConaughey, and R. Shekhar, "Automated identification of innocent still's murmur in children," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 6, pp. 1326–1334, Jun. 2017.
- [22] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [23] H. Lee *et al.*, "Fully automated deep learning system for bone age assessment," *J. Digit. Imag.*, vol. 30, no. 4, pp. 427–441, Aug. 2017.
- [24] J. Li *et al.*, "Automatic classification of fetal heart rate based on convolutional neural network," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1394–1401, Apr. 2019.
- [25] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, p. 2181, 2016.
- [26] Y. Ma, Y. Cao, S. Vrudhula, and J.-S. Seo, "Performance modeling for CNN inference accelerators on FPGA," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, to be published.
- [27] M. E. Markaki, I. Germanakis, and Y. Stylianou, "Automatic classification of systolic heart murmurs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2013, pp. 1301–1305.
- [28] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation—An event detection approach using deep recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, Sep. 2018.
- [29] M. Mishra, H. Menon, and A. Mukherjee, "Characterization of S₁ and S₂ heart sounds using stacked autoencoder and convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3211–3220, Sep. 2019.
- [30] T. Nilanon, J. Yao, J. Hao, S. Purushotham, and Y. Liu, "Normal/abnormal heart sound recordings classification using convolutional neural network," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 585–588.
- [31] S. A. Osia *et al.*, "A hybrid deep learning architecture for privacy-preserving mobile analytics," *arXiv preprint arXiv:1703.02952*, 2017.
- [32] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1138–1152, Jul. 2014.
- [33] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nat. Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, Mar. 2018.
- [34] 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 *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 621–624.
- [35] E. Pretorius, M. L. Cronje, and O. Strydom, "Development of a pediatric cardiac computer aided auscultation decision support system," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6078–6082.
- [36] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Domínguez, "Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals," *Ann. Biomed. Eng.*, vol. 38, no. 1, pp. 118–137, Jan. 2010.
- [37] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sriraman, "Classifying heart sound recordings using deep convolutional neural networks and Mel-frequency Cepstral coefficients," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 813–816.
- [38] H. Ryu, J. Park, and H. Shin, "Classification of heart sound recordings using convolution neural network," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 1153–1156.
- [39] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. A. Ramaiah, "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1407–1414, 2013.
- [40] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4510–4520.
- [41] R. SaracOğlu, "Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction," *Eng. Appl. Artif. Intell.*, vol. 25, no. 7, pp. 1523–1528, 2012.
- [42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [43] A. A. Sepehri, A. Kocharian, A. Janani, and A. Gharebaghi, "An intelligent phonocardiography for automated screening of pediatric heart diseases," *J. Med. Syst.*, vol. 40, no. 1, p. 16, Oct. 2015.
- [44] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
- [45] S. Singh, Y.-S. Jeong, and J. H. Park, "A survey on cloud computing security: Issues, threats, and solutions," *J. Netw. Comput. Appl.*, vol. 75, pp. 200–222, Nov. 2016.
- [46] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Apr. 2016.
- [47] N. Strodthoff and C. Strodthoff, "Detecting and interpreting myocardial infarctions using fully convolutional neural networks," *Physiol. Meas.*, vol. 40, no. 1, Jan. 2019, Art. no. 015001.
- [48] S. Sun, Z. Jiang, H. Wang, and Y. Fang, "Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 219–230, 2014.
- [49] H. Tang, T. Li, T. Qiu, and Y. Park, "Segmentation of heart sounds based on dynamic clustering," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 509–516, 2012.
- [50] C. Thomae and A. Dominik, "Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 625–628.
- [51] H. Uğuz, "A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases," *J. Med. Syst.*, vol. 36, no. 1, pp. 61–72, Feb. 2012.
- [52] Y. Xu, B. Xiao, X. Bi, W. Li, J. Zhang, and X. Ma, "Pay more attention with fewer parameters: A novel 1-D convolutional neural network for heart sounds classification," in *Proc. Comput. Cardiol. Conf. (CinC)*, vol. 45, Sep. 2018, pp. 1–4.
- [53] Y. Yang, Z. Zhong, T. Shen, and Z. Lin, "Convolutional neural networks with alternately updated clique," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2413–2422.
- [54] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," *Expert Syst. Appl.*, vol. 84, pp. 220–231, Oct. 2017.
- [55] Y. Zhang, L. Sun, H. Song, and X. Cao, "Ubiquitous WSN for healthcare: Recent advances and future prospects," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 311–318, Aug. 2014.

Bin Xiao received the B.S. and M.S. degrees in electrical engineering from Shaanxi Normal University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Xidian University, Xi'an.

He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include image processing and pattern recognition.

Yunqiu Xu received the B.E. and M.E. degrees from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2015 and 2019, respectively.

His research interests include pattern recognition, deep learning, and biomedical signal processing.

Xiuli Bi received the B.Sc. and M.Sc. degrees from Shannxi Normal University, Xi'an, China, in 1995 and 1998, respectively, and the Ph.D. degree in computer science from the University of Macau, Macau, China, in 2017.

She is currently an Associate Professor with the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China. Her research interests include image processing, multimedia security, and image forensics.

Weisheng Li received the B.S. and M.S. degrees from the School of Electronics and Mechanical Engineering, Xidian University, Xi'an, China, in 1997 and 2000, respectively, and the Ph.D. degree from the School of Computer Science and Technology, Xidian University, in 2004.

He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research focuses on intelligent information processing and pattern recognition.

Zhuo Ma received the Ph.D. degree in computer architecture from Xidian University, Xi'an, China, in 2010.

He is currently an Associate Professor with the School of Cyber Engineering, Xidian University, Xi'an. His research interests include cryptography, machine learning in cyber security, and Internet of Things security.

Junhui Zhang received the B.S. degree in clinical medicine from Sichuan Luzhou Medical College, Luzhou, China, in 1995, and the M.S. and Ph.D. degrees in obstetrics and gynecology from Chongqing Medical University, Chongqing, China, in 2008 and 2015, respectively.

She is currently a Professor with the Hospital of Chongqing Medical University, Chongqing, and an Associate Chief Physician with the Health Management Center, the First Affiliated Hospital of Chongqing Medical University, Chongqing. Her research interests include gynecologic endocrine and fetal dysplasia.

Xu Ma received the master's degree in genetics from the West China University of Medical Science, Chengdu, China.

He is with the Human Genetics Resource Center, National Health Commission, Beijing, China, and a Professor with Peking Union Medical College, Beijing. He had published over 280 papers. His research interests include genetics, epidemiology, and statistics.