LEPCNet: A Lightweight End-to-End PCG Classification Neural Network Model for Wearable Devices

Lixian Zhu, Wanyong Qiu, Yu Ma, Fuze Tian, Mengkai Sun, Zhihua Wang, Kun Qian, *Senior Member, IEEE*, Bin Hu, *Fellow, IEEE*, Yoshiharu Yamamoto, *Member, IEEE*, and Björn W. Schuller, *Fellow, IEEE*

***Abstract*— Wearable intelligent phonocardiogram (PCG) sensors provide a noninvasive method for long-term monitoring of cardiac status, which is crucial for the early detection of cardiovascular diseases (CVDs). As one of the key technologies for intelligent PCG sensors, PCG classification techniques based on computer audition (CA) have been widely leveraged in recent years, such as convolutional neural networks (CNNs), generative adversarial nets, and long short-term memory (LSTM). However, the limitation of these methods is that the models have a sizeable computational complexity, which is not suitable for wearable devices. To this end, we propose an end-to-end neural network for PCG classification with low-computational complexity [52.67k parameters and 1.59M floating point operations per second (FLOPs)]. We utilize two public datasets to test the model, and experimental results demonstrate that the proposed model achieves an accuracy of 93.1% in the 2016 PhysioNet/CinC Challenge 2016 dataset with considerable complexity reduction compared with the state-of-the-art works. Moreover, we design an energy-efficient wearable PCG sensor and deploy the proposed algorithms on it. The experimental results show that our proposed model consumes only 245.1 mW for PCG classification with an accuracy of 89.8% on test datasets. This means that the proposed model obtains excellent performance compared with previous work while consuming lower power, which is significant in practical application scenarios.**

***Index Terms*— Computer audition (CA), end to end, heart sound, intelligent sensor, lightweight model.**

This work was supported in part by the Ministry of Science and Technology of the People's Republic of China within the STI2030- Major Projects under Grant 2021ZD0201900; in part by the National Natural Science Foundation of China under Grant 62227807 and Grant 62272044; in part by the Teli Young Fellow Program from the Beijing Institute of Technology (BIT), China; in part by the Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) under Grant 20H00569; in part by the BIT Research and Innovation Promoting Project under Grant 2022YCXZ012; and in part by the China Postdoctoral Science Foundation under Grant 2023M730250. The Associate Editor coordinating the review process was Dr. Octavian Postolache. *(Corresponding authors: Kun Qian; Bin Hu.)*

Lixian Zhu, Wanyong Qiu, Yu Ma, Fuze Tian, Mengkai Sun, Kun Qian, and Bin Hu are with the Key Laboratory of Brain Health Intelligent Evaluation and Intervention, Ministry of Education, and the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: zhulx17@bit.edu.cn; qiuwy@bit.edu.cn; mayu@bit.edu.cn; tianfz@bit.edu.cn; smk@bit.edu.cn; qian@bit.edu.cn; bh@bit.edu.cn).

Zhihua Wang and Yoshiharu Yamamoto are with the Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Tokyo 113-0033, Japan (e-mail: wzhihua@p.u-tokyo.ac.jp; yamamoto@p.u-tokyo.ac.jp).

Björn W. Schuller is with the Group on Language, Audio, and Music (GLAM), Imperial College London, SW7 2AZ London, U.K., and also with the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, 86159 Augsburg, Germany (e-mail: schuller@ieee.org).

Digital Object Identifier 10.1109/TIM.2023.3315401

I. Introduction

AS ONE of the leading killers, cardiovascular diseases (CVDs) have made a large number of deaths [1]. The early detection of CVDs is essential to reduce mortality. Heart sound, a reflection of cardiac activity, is often utilized to detect CVDs. In most cases, clinicians diagnose CVDs through a stethoscope in a few minutes. However, clinician performance and access to service vary widely, especially in low-income countries and areas with poor medical equipment [2]. The patient's treatment will be delayed if clinicians cannot quickly locate the abnormal heart activity. What is worse, it is difficult for patients to monitor their daily heart activities to the extent they have to go to the hospital for diagnosis. Unfortunately, abnormal heart sounds are difficult to detect in the short term, especially in the early stages of heart disease when the symptoms are not obvious. This is because the randomness and variability of CVDs symptoms cause complexity and diversity of heart sound signals [3]. Therefore, long-term monitoring and identification of heart sounds are desirable.

The phonocardiogram (PCG) is a visual waveform of heart sounds, which makes it easy to extract features for heart sound classification. With the development of computer audition (CA), PCG-based classification methods have been widely introduced [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. For these methods, most research efforts have focused on feature-based classification methods. A multitude of feature extraction techniques has been applied to algorithms based on machine learning (ML) and deep learning (DL) [2], [15], [16], [17], [18], [19], such as wavelet coefficients [17], statistical features [10], mel-frequency cepstral coefficients (MFCCs) [13], short-time Fourier transform (STFT) [15], spectrograms [19], and others. Although feature engineering partially achieves higher classification accuracy, its limitations are also evident. First, feature engineering usually requires the extraction of time–frequency features, which greatly increases the computational complexity and power consumption of the model. Second, the type of features affects the classification results of the model, which is highly dependent on human knowledge and experience. In addition, to achieve higher accuracy, researchers usually employ two types of methods,

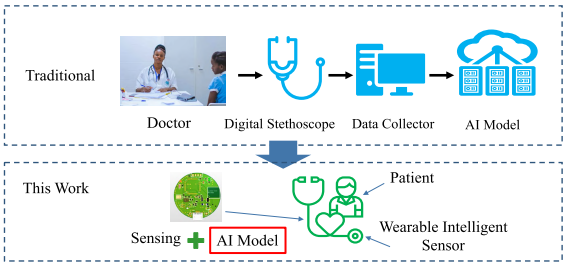


Fig. 1. Proposed PCG sensor with an integrated lightweight AI model.

one adopting a multifeature fusion strategy and the other using a larger or deeper network structure. Nevertheless, these approaches, both of which require highly capable computers, are difficult to deploy on wearable devices, so that monitoring CVDs in real time is not possible. Intelligent PCG sensors embedded with a lightweight model can address this issue.

In addition, a series of wearable PCG devices have been proposed in the previous studies [20], [21], [22], [23], [24]. Although these devices can capture heart sound signals, they do not feature integrated intelligent PCG classification algorithms to perform abnormal heart sound detection. There are typically two strategies to deploy these algorithms. One is to implement them through the graphics processing unit (GPU), the field-programmable gate array (FPGA), or an application specific integrated circuit (ASIC) [25], [26], [27], [28], [29]. These devices can effectively accelerate neural networks. However, the power consumption is high, and they are not suitable for integration into wearable devices. Another strategy is to use the microcontroller unit (MCU) to design an electronic stethoscope that would then wirelessly transmit data to a computer or smart mobile device for analysis [30]. However, this would consume more time and would not be suitable for low-income countries and regions.

Consequently, as illustrated in Fig. 1, we propose a new method in this study to address these disadvantages of state-of-the-art works. The specific framework is shown in Fig. 2. First, to deploy the PCG classification algorithm, we design and implement a digital stethoscope with extremely low-power consumption while operating. Second, a lightweight end-to-end neural network model for PCG classification is proposed to reduce computational power and complexity while maintaining high accuracy in hardware deployment.

The rest of the work is organized as follows. Previous studies related to PCG classification implemented and deployed using different methods are reviewed in Section II. Section III describes the proposed lightweight neural network model. Section IV shows the implementation hardware details of PCG sensors. The performance of the proposed model is demonstrated in Section V. In Section VI, we discuss the results by comparing the proposed method with state-of-the-art works. Finally, we conclude in Section VII.

II. Related Work

Numerous research efforts on PCG classification utilizing ML and DL have been reported in the previous works over the years. To materialize PCG classification, researchers have applied several ML algorithms, such as support vector machines (SVMs) [31], [32], k-nearest neighbors (k-NNs) [33], random forest [34], or hidden Markov models [35]. These methods conjoin signal processing and feature extraction techniques, including time-domain features, frequency-domain features, and time–frequency features, to achieve the anomaly classification of PCG. However, they are not well generalizable, as most methods rely on manually engineered features. As DL develops, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their multiple derived and related versions are widely utilized for PCG classification to solve both the accuracy concern and generalization problems [36]. With sophisticated network frameworks and powerful self-learning, DL algorithms achieve higher accuracy, yet they simultaneously require considerable computation and power consumption, which makes them unfit to be deployed in wearable devices or mobile platforms. Recently, lightweight classification methods have been proposed, and one of these strategies is to change the architecture of the network based on feature engineering to reduce model complexity. For instance, Munia et al. [10] segmented the PCG raw signal into slices with 1.5-s length and then extracted STFT features and fed them into a lightweight CNN to improve accuracy. Another strategy is to design an end-to-end classification method, where the PCG signal can be directly fed into the DL network after simple processing and segmentation to learn features instead of extracting them manually. Shuvo et al. [4] proposed a novel lightweight end-to-end convolutional RNN (CRNN) architecture for heart disease, which has three representation learning modules to learn the features of the signal. Tian et al. [37] performed two-stage training using the original PCG signal to alleviate the limitations of imbalance in datasets, which significantly reduced the computational load compared with ResNet [38]. Xiao et al. [39] employed a sliding window to segment the PCG signal and then fed it into a deep CNN architecture with an attention mechanism, which reduced the number of parameters compared with state-of-the-art methods.

Of these methods, most studies have focused on anomaly classification rather than on predicting heart disease directly from heart sounds. Only a few works have centered on multiclass classification (normal, artifact, murmurs, and extrasystole) [4], [40], [41]. One reason for this is the paucity of available multiclass datasets, and another is the severe imbalance of available datasets, such as the PhysioNet/CinC Challenge 2016 dataset [42].

Another important aspect of wearable intelligent PCG sensors is PCG classification artificial intelligence (AI) algorithm deployment in hardware. In the past decade, FPGA, GPU, and ASIC have been widely used for the acceleration of neural networks in edge computing [25]. For PCG classification, however, researchers mainly leveraged FPGA to implement AI algorithms. Jhong et al. [28] proposed a PCG multiclass classification method with a DL hardware/software codesign (Zynq platform), the processing system (PS) for training and feature extraction, and programmable logic (PL) for model deployment. Li et al. [29] proposed an acceleration scheme for heart sound classification based on a system on a chip-field programmable gate array (SOC-FPGA), such that the

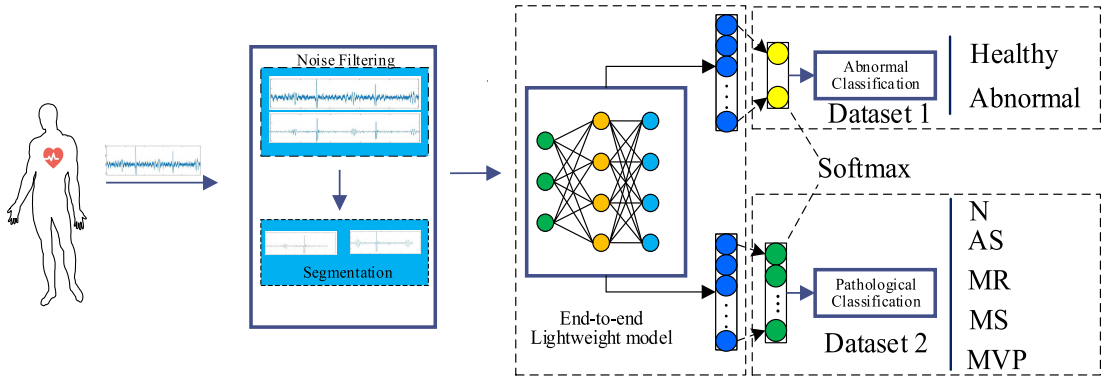


Fig. 2. System architecture of the proposed wearable intelligent PCG monitoring. After several generic preprocessing steps, the heart sound signals are fed into the proposed lightweight model to carry out a two-dataset classification of CVDs: 1) abnormal classification and 2) pathological (normal (N), MR, MS, MVP, and AS) classification.

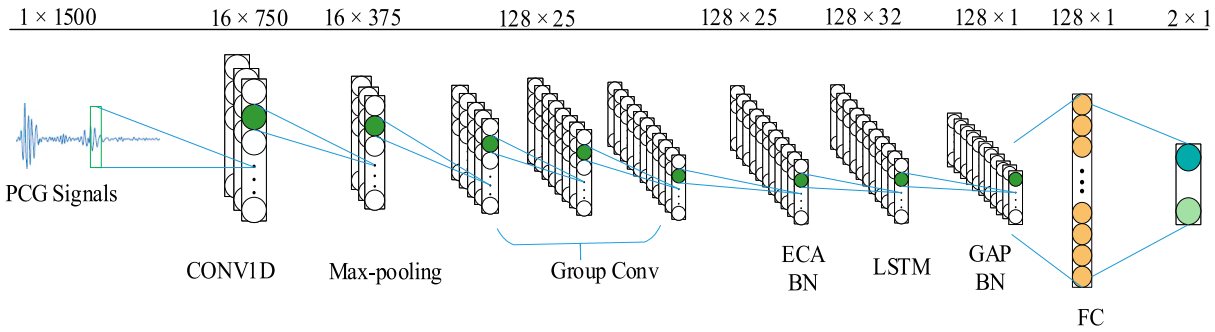


Fig. 3. Detailed architecture of the proposed lightweight model.

speed of classification is 3.13 times faster than the central processing unit (CPU). Clifford et al. [43] implemented a discrete wavelet transform (DWT) algorithm with Shannon energy on an FPGA for real-time PCG diagnosis. In addition, Son and Kwon [40] designed a cardiac auscultation monitoring system based on wireless sensing, such that the mobile phone and cloud platform are used to deploy the classification model. Although the computation efficiency has been significantly improved by using an FPGA or the GPU [26], [27], the power consumption is not negligible, making it a challenge for wearable intelligent PCG devices. From a practically applicable point of view, an MCU is a better choice and has been widely utilized in commercial digital stethoscopes. Nevertheless, how to make the MCU run the neural networks is a challenge, and these products generally use the MCU to acquire heart sounds for further analysis by transmitting them to the PC or mobile phone.

III. Proposed Lightweight End-to-End PCG Classification Model

A. Proposed Lightweight Neural Network

To be suitable for low-power embedded devices, we propose a lightweight end-to-end neural network model with the architecture shown in Fig. 3. The first layer is a 32 convolutional layer with a 3 × 3 kernel followed by a 2 × 2 max-pooling layer. The group convolutional layers are stacked over the first layer, as described in Section III-B. After the group convolutional layer, the effective attention module (EcaNet) [44], a channel attention mechanism proven effective in improving the efficiency of convolutional networks, is applied to capture cross-channel information. EcaNet provides a novel way of increasing channel correlation and reducing dimensionality without losing information, as described in [45]. Considering the importance of temporal features in the PCG as well, long short-term memory (LSTM) is embedded in the neural network to capture more time-domain features. Finally, the outputs of all these layers are flattened to connect a fully connected (FC) layer to predict each probability with a SoftMax layer. In addition, a rectified linear unit (ReLU) activation layer is applied to the first convolutional and FC layers to increase nonlinearities in the computation and reduce convergence times. Max pooling is embedded in the group convolution to reduce the dimensionality of the feature map. Max pooling, global average pooling (GAP), and group convolution work together to achieve lightweight goals. To overcome the problem of imbalanced data, a dropout layer follows each convolutional module, and the FC layer connects the batch normalization (BN) layer. Also, $L2$ regularization is applied to the SoftMax layer to prevent overfitting and improve the robustness of the model. The adaptive learning rate optimizer (Adam) is used to compile the model with a learning rate of 0.0008.

B. Proposed GCT

One of the challenges of deploying CNN-based PCG classification algorithms on low-power mobile devices is related to their enormous computing operations, such as additions and multiplications, mainly attributed to the traditional convolutional layer. Inspired by Ma et al. [46], we propose a group convolution technique (GCT) based on depthwise separable convolution instead of traditional convolutional layers to reduce the operations.

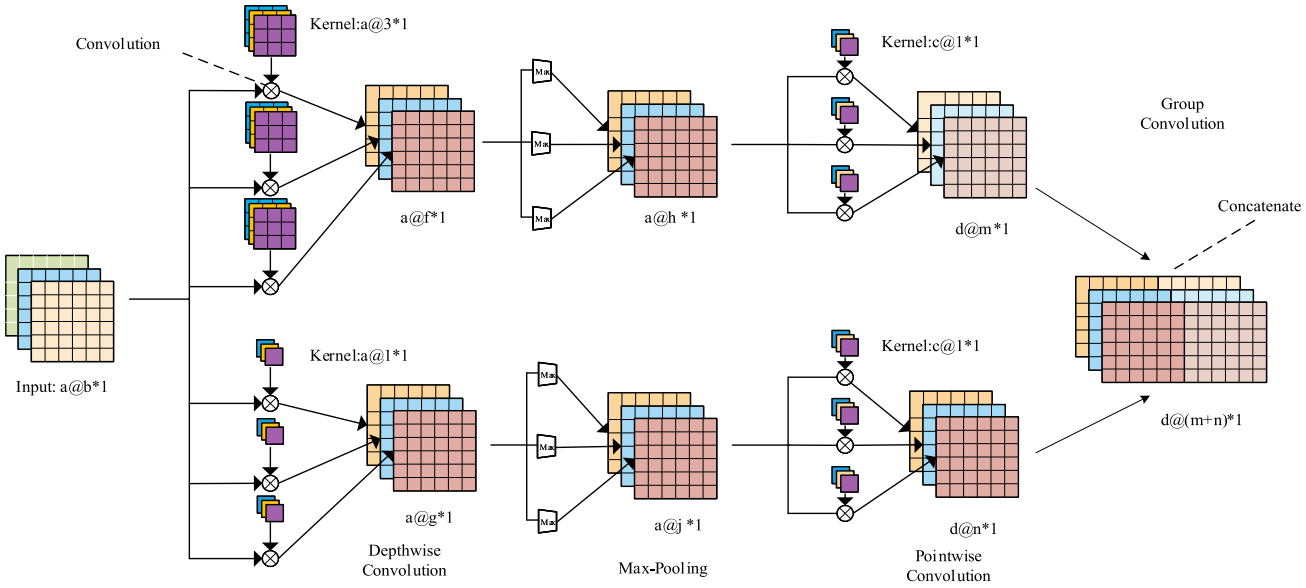


Fig. 4. Proposed GCT ("*a*@*b***c*": the number of input channels is "*a*," and the input size is "*b***c*").

As shown in Fig. 4, there are four components comprising group convolution, namely, depthwise convolutional layer, max-pooling layer, pointwise convolutional layer, and concatenate layer. The depthwise convolutional layer and pointwise convolution layer conjointly form depthwise separable convolution, which was first used by Howard et al. [47] for image classification and proved to be able to reduce the number of operations dramatically. Likewise, the max-pooling layer reduces the computation by eliminating nonmaximal values. The concatenate layer is used to merge the data from the two groups of depthwise separable convolution data as an output feature map.

The max-pooling layer, which commonly follows after the convolutional layer, reduces the dimensionality of the feature map while increasing the robustness of the model. Furthermore, the pointwise convolution is a special form of conventional convolution at a kernel size of 1×1. It means that pointwise convolution also bears redundant information requiring compression. In light of this, we modify the position of max pooling by moving it between the depthwise convolution layer and the pointwise convolution layer. As shown in Fig. 4, the group convolution obviously reduces the parameters and operations compared with traditional convolution.

As shown in Table I, for the GCT, the number of parameters and computation of the model are compressed, and its compression ratios are the following equations relative to the traditional convolution:

$$\mathrm{RP}_P = \frac{1}{k_1} + \frac{1}{ak_1} + \frac{1}{ab} \tag{1}$$

$$\mathrm{RP}_F = \frac{b-1}{k_1 b} + \frac{1}{2b} \tag{2}$$

where RP_P and RP_F are denoted as the compression ratios of the parameters and operations in the model, respectively. a is the number of input channels for depthwise convolution, and b is the number of output channels for pointwise convolution. In addition, k corresponds to the kernel length of the depthwise convolution, h represents the output feature maps length of the pointwise convolution, and w denotes the output feature maps width of the pointwise convolution. The above equations are based on the conditions $k2 = 1$, $h1 = h2 = 1/2h$, and $w1 = w2 = 1/2w$. As an example, for $k1 = 3$, $b = 64$, and $a = 16$, the parameters and floating point operations per second (FLOPs) of group convolution are only 38.1% and 11.7% of those of conventional convolution, respectively.

TABLE I
COMPARISON OF COMPUTATIONS BETWEEN CONVENTIONAL AND GCTS

	Parameters	**FLOPs**
Conventional	$k_1 \times 1 \times a \times b$	$k_1 \times 1 \times a \times b \times h \times w$
GCT	$a \times (b \times k_1 + k_2)$	$b \times ((k_1 + a) \times h_1 \times w_1 + (k_2 + a) \times h_2 \times w_2)$

Note: GCT has two depthwise-and-pointwise convolutions, and k_1, h_1, w_1 denote the first depthwise-and-pointwise convolution parameters, k_2, h_2, w_2 denote the second depthwise-and-pointwise convolution parameters, respectively. We assume that one of the GCT convolutions has the same size as the conventional convolution in terms of kernel and the output feature map.

C. BCE Loss

As aforementioned, there is an imbalance in the number of abnormal patients and normal subjects in dataset 1, which affects the classification accuracy of the model. A new loss function, named balanced cross entropy (BCE), is, therefore, introduced to address the shortcoming of an imbalanced dataset. We modify the loss function based on the standard cross-entropy loss, with the standard cross-entropy loss $L_1(p, q)$ formula as in (3) and the BCE loss $L_2(p, q)$ function formula as in (4)

$$L_1(p, q) = -\sum_{i=1}^{C} p_i \log q_i \tag{3}$$

$$L_2(p, q) = -\frac{\beta}{1-\beta} \sum_{i=1}^{C} p_i \log q_i \tag{4}$$

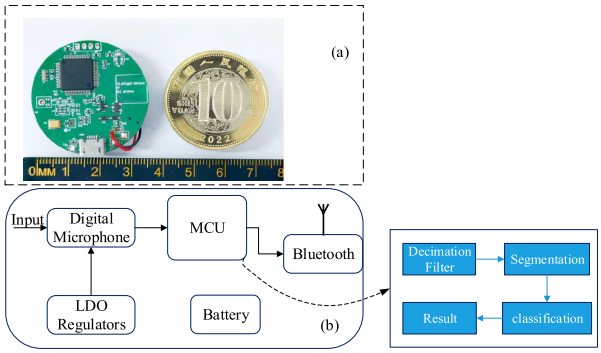


Fig. 5. Details of the wearable PCG sensor. (a) Appearance of the sensor. (b) Internal structure of the wearable PCG sensor.

TABLE II
SPECIFICATION OF THE HEART SOUND SENSOR

PCG Senor	
Voltage VCC	3.7 V
Standby Current	35 mA
Sample Rate	1Hz-3000 Hz
Sensitive	94dB SPL @ 1 kHz
Output Amplitude	0∼2.5 V
Battery	300 mAh

where p_i is the probability distribution of the true sample, q_i is the predicted sample probability distribution as $q_i = (\exp(z_y))/(\sum_{y=1}^{k} \exp(z_y))$, z_y denotes the predicted output of class y in the model, and C is the total number of classes.

For binary classification, 1 and -1, the BCE weighting factor $\beta \in [0, 1)$ corresponding to class 1, and $1 - \beta$ corresponding to class -1. β is given by the following equation:

$$\beta = \frac{n}{m} \tag{5}$$

where n is the number of samples for class 1 and m is the number of total training samples. In this work, we set the β to 0.25 to overcome the dataset imbalance issue.

IV. Wearable Intelligent Sensors

In this work, we propose a lightweight neural network for heart sound classification that can be deployed on low-power embedded devices. It reduces development costs and enables flexible updating of the model compared with implementation on the ASIC or FPGA. We also design an intelligent PCG sensor (as shown in Fig. 5) to test and optimize the proposed model. It plays a fundamental role in the intelligent monitoring of CVDs. Table II shows some implementation details of the PCG sensors. The component of microelectromechanical systems (MEMSs), which is a digital microphone for heart sounds acquisition, provides the input to the PCG sensor. The PCG sensor contains an MCU named STM32F405RGT6, which performs preprocessing for PCG data and runs the neural network model. The core of the MCU is an ARM 32-bit Cortex-M4 CPU with a floating-point unit (FPU), which is used for data processing. The MCU also contains 1-MB flash and 192-kB SRAM, where the flash stores code and SRAM stores variables, such as the feature map of the model. Moreover, a lithium battery with a 300-mAh capacity powers the sensor, and the low dropout (LDO) regulates the battery output to a stable 2.5 V for the microphone and the MCU.

TABLE III
DETAILS OF DATASET 1

Subset	Total	Abnormal Recordings	Normal Recordings
Traning-a	409	292	117
Traning-b	490	104	386
Traning-c	31	24	7
Traning-d	55	28	27
Traning-e	2141	183	1958
Traning-f	114	34	80
Total	3240	665	2575

V. Experimental Results

A. Experimental Setup

The proposed model is trained on a GTX3080 GPU using PyTorch and Keras. The model training parameters are set to an adaptive learning rate optimizer (Adam) with a learning rate of 0.0008. Also, a batch size of 32 is used for training and validation. Furthermore, both datasets 1 and 2 are trained using the tenfold cross-validation method, and the experimental results are given in the form of mean and standard deviation.

B. Dataset and Preprocessing

In this work, we choose two public datasets to evaluate the performance of the proposed model. The first is the PhysioNet/CinC Challenge 2016 dataset, which consists of six subdatasets with 3240 heart sound recordings (training-a, training-b, training-c, training-d, training-e, and training-f, collected by six different research groups and details, are shown in Table III). The recording environment for this dataset includes clinical and nonclinical and, hence, inevitably varying levels of noise. All data are in a uniform ".wav" format, with a sampling rate of 2 kHz and lengths ranging from 5 to 150 s. The dataset is comprised of heart sound data from healthy subjects and pathological patients, yet is only annotated with a binary class (normal and abnormal), and not with pathological cases, such as mitral regurgitation (MR), aortic stenosis (AS), valvular stenosis (VS), and so on. To validate the robustness of the proposed model, the dataset in [40] is also utilized as a secondary dataset. The dataset contains a total of 1000 PCG recordings with five classes, i.e., 200 normal (N), and 800 abnormal containing 200 MR, 200 AS, 200 mitral stenosis (MS), and 200 mitral valve prolapse (MVP). Each PCG recording is approximately 3 s in length with an 8-kHz sampling rate and contains three heart sound cycles [40]. Fig. 6 shows a waveform for each class.

Considering the diversity of conditions to collect datasets, it will inevitably make the data contain various noises. A preprocessing stage is, therefore, essential. In this work, all the data are resampled to 1000 Hz and filtered by a third-order Butterworth bandpass filter with a cutoff frequency of 20–400 Hz to filter out noise. The PCG recordings are segmented into 1.5-s lengths by Springer et al. [53] for further analysis.

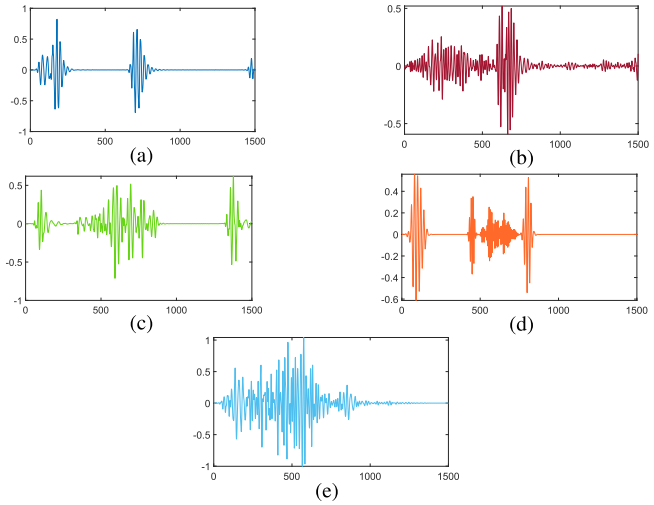


Fig. 6. Waveform of the existing five CVD classes in dataset 2. (a) Normal. (b) MR. (c) MS. (d) MVP. (e) AS.

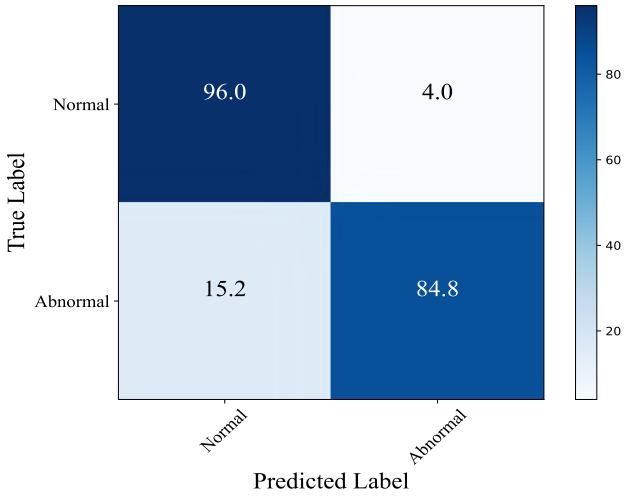


Fig. 7. Normalized confusion matrix (in [%]) for dataset 1. The positive and negative sample sizes are 2817 and 749, respectively.

C. Comparison With Other Works

1) MobileNetV3-Small: MobileNetV3 [54] is a classic and increasingly utilized lightweight deep neural network. Its bottleneck inherits the depth separable convolution of MobileNetV1 and the inverted residual of MobileNetV2. MobileNetV3-Small is a special form of MobileNetV3 with 11 blocks and four convolutional layers. To demonstrate the excellence of the proposed model, we keep the original network structure unchanged and only adjust the convolutional kernel from 2-D to 1-D.

2) ShuffleNetV2: The goal of ShuffleNetV2 [46] is to speed up the run time of the network while lowering the number of parameters. It has an efficient network structure based on ShuffleNetV1 with the addition of a channel split. Only the convolutional kernel size is modified to match the input features in this work.

3) ResNet: ResNet is one of the most popular DL frameworks and has been widely featured in various baselines. It aims to solve the problem of gradient disappearance as the depth of a neural network increases. For a fair comparison, we modify the kernel size of the convolutional layers.

4) Xception: Xception is an improved form of inception that is entirely decoupled from cross-channel correlations and spatial correlations in neural network feature maps [55]. We only modify the kernel size of the convolutional layers.

D. Evaluation Metrics

In our experiments, we utilize two datasets to perform the proposed model, and the details of these results are explained and discussed in Sections V-D, V-E, and VI.

We first go through the calculation of the experimental metrics. As shown in Table IV, TP means true positive, TN is true negative, FP is false positive, and FN denotes false negative. Furthermore, TP and TN represent the correct identification of the PCG, respectively. FP and FN indicate an incorrect classification of PCG, respectively. To determine results for the classification of PCG, we use four evaluation metrics, which are accuracy (Acc), sensitivity (Sen), specificity (Spec), and positive precision (Pre). Their formulas are as follows, respectively:

TABLE IV
DEFINITION OF METRICS CALCULATION

True \ Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{8}$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{9}$$

Note that sen increases with pre and vice versa. Thus, we also calculate the $F1$ score to evaluate the model performance; it is calculated by the following equation:

$$F1 \text{ score} = \frac{2 \times \text{pre} \times \text{sen}}{\text{pre} + \text{sen}}. \tag{10}$$

1) Performance on Dataset 1: Table V shows the experimental results of our proposed model for PCG classification and comparison with previous work. For fair comparisons, we utilized the same dataset as the state-of-the-art work. It can be observed that the proposed model achieves superior classification performance in terms of accuracy, sensitivity, precision, and $F1$ score (as shown in Fig. 7).

2) Performance Dataset 2: To demonstrate the performance of the proposed model, pathological classification is performed on the secondary dataset. Table VI shows the experimental results for five-class pathological classification. Clearly, the proposed model yields near perfect results, with an accuracy and an $F1$ score of 99.4% and 99.0%, respectively. Also, the proposed model achieves better performance, as is evident in Fig. 8 (confusion matrix).

TABLE V

PERFORMANCE COMPARISON BETWEEN THE PROPOSED ALGORITHM AND PREVIOUS WORKS IN DATASET 1 [%]

	[9]	[48]	[7]	[49]	[4]	Baseline	This Work
Method	1D-CNN	2D-CNN	1DCNN+2DCNN	1DCNN	2DCNN + LSTM	-	1DCNN+LSTM
Acc	82.2	86.2	89.2	-	86.6	90.2	**93.1±1.1**
Sen	90.6	-	89.9	90.9	89.8*	91.9	**88.9±1.9**
Spec	71.2	-	86.4	83.3	84.4*	87.4	**90.4±1.5**
Macc	80.9#	85.1	88.1	87.1	87.1*	89.1	**89.7±1.6**
F1-score	81.9	84.1	88.1#	86.9	87.1*	89.1	**89.4±1.6**

*: calculated from the confusion matrix.
#: calculated from equations (7), (8), (9), and (10).
Note: Macc = (Spec + Sen) / 2. The metrics for [9], [48], [7], [49] and [4] are all derived from the original literature.

TABLE VI

PERFORMANCE COMPARISON BETWEEN THE PROPOSED ALGORITHM AND PREVIOUS WORKS IN DATASET 2 [%]

	[40]	[41]	[25]	[50]	[51]	[52]	This Work
Method	DNN	Random Forest	MCC	Deep Wavenet	CNN	CNN-BiLSTM	1DCNN+LSTM
Acc	97.9	95.1	98.3	97.0	98.6	99.3	**99.4±0.5**
Sen	-	-	-	92.5	98.3*	98.3	**99.1±0.7**
Spec	-	-	-	98.1	-	99.6	**98.9±0.4**
F1-score	99.7	-	-	95.2	98.5*	98.3	**99.0±0.5**

*: Calculated from the average of the experimental results in [51]
Note: The metrics for [40], [41], [25], [50], [51] and [52] are all derived from the original literature.

TABLE VII

COMPARISON OF COMPUTATIONAL COMPLEXITY WITH THE STATE-OF-THE-ART METHODS

	[9]	[48]	[7]	[49]	[13]	[56]	Baseline	Proposed
FLOPs	999622*	138.85M*	3824000*	6614272 *	4476728*	521214393*	899659	**1598971**
Parameters	200122*	785904*	570440*	199376*	4.8M*	19408911*	179438	**52677**
Method	1D-CNN	2D-CNN	1D-CNN+2D-CNN	1D-CNN	2D-CNN+RNN	2D-CNN	-	1D-CNN+LSTM
Input Size	1×2500	-	1×1000	1×2500	499×39	129×129	-	1×1500
Conv Layers	-	5	6	9	3	2	-	4
FC layers	-	1	1	1	1	3	-	1
Macc(%)	81.5	85.1	88.2	87.1	97.3#	-	88.8	**89.7**

*: The values are estimated based on the method structure provided in [9] [48] [7] [49] [13] [56].
#: Calculated from the equation: macc = (spec + sen) / 2.

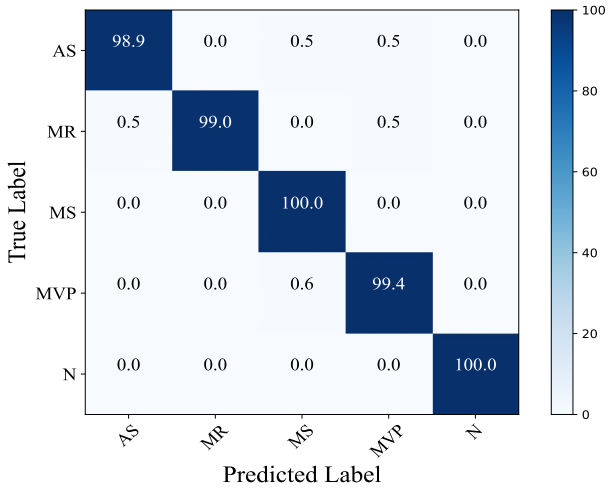


Fig. 8. Normalized confusion matrix (in [%]) for dataset 2. The sample sizes for the five classes (AS, MR, MS, MVP, and N) are 187, 196, 173, 163, and 180, respectively.

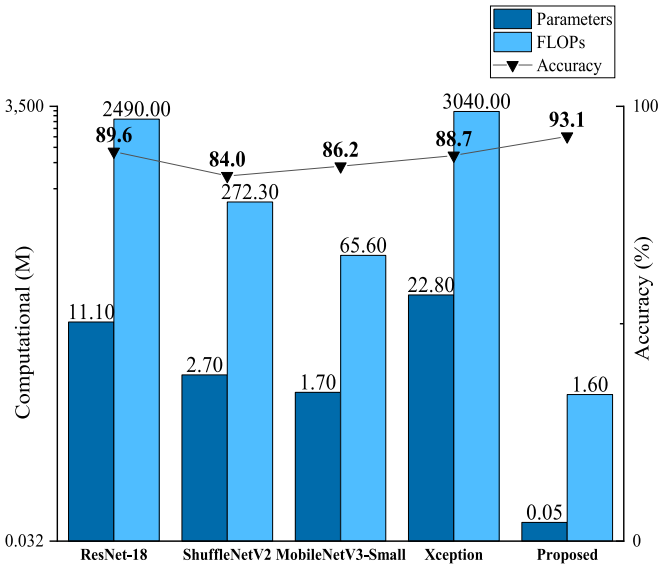


Fig. 9. Performance comparison between the proposed model and classic DL models.

3) Computational Performance as a Lightweight Neural Network: Fig. 9 provides a clear comparison of our proposed lightweight model and other state-of-the-art lightweight models, such as MobileNetV3-Small, ShuffleNetV2, ResNet-18, and Xception. In terms of accuracy, our proposed model outperforms the popular methods, including ResNet-18, ShuffleNetV2, MobileNetV3-Small, and Xception relatively by

TABLE VIII
EXPERIMENTAL RESULTS OF THE PROPOSED MODEL IN HARDWARE DEPLOYMENT

	Time(ms)		Power(mW)		Accuracy(%)	
	PC	PCG sensor	Standby	Dynamic	Original	PCG sensor
Dataset1	10.1	360161.0	129.7	244.6	92.9	89.8
Dataset2	2.0	90589.0	130.2	245.7	99.4	95.6

Note: The dataset 1 has 200 test samples and dataset 2 has 50 test samples.

TABLE IX
MINIMUM RESOURCES REQUIRED FOR DIFFERENT MODELS IN THE PCG SENSOR

	ResNet	ShuffleNet	MobileNet	Xception	Proposed
ROM(MB)	42.62	10.35	7.58	79.25	120.00
RAM(KB)	942.00	474.44	289.69	4.35	59.04
MACC(M)	2375.93	258.71	63.43	2896.23	1.60

3.5%, 9.1%, 6.9%, and 4.4%, respectively, while the number of parameters and amount of operations are significantly reduced, thereby further accelerating the running time of the model and reducing power consumption. This makes our proposed lightweight model better applicable to low-resource wearable devices for real-time CVDs' monitoring.

Table VII compares the complexity of the proposed method with the baseline and several state-of-the-art works. Considering these methods, we design the baseline metrics, as shown in Tables V and VI, which are designed with the rule that the maximum value of each row in the table plus 1% is used as the baseline metrics, with the purpose of illustrating the excellent performance of the model proposed in this article. Also, the setting rule for the baseline in Table VII is to calculate the average of the metrics corresponding to the other works in Table VII. We design the lightweight network with this as our goal. Obviously, the proposed model, with the exception of sensitivity, has a remarkable improvement in all of the metrics, indicating that the proposed method is feasible, as the details will be shown in the discussion section.

E. Real-Time Execution

Smart wearable devices have a smaller size and lower power consumption compared with traditional computers. As a result, wearable devices need to meet continuous operation and be designed with classification models that have low-computing intensity. As can be seen from Table VIII, the method proposed in this work, which is obviously less computationally intensive, is deployed on the PCG sensor to further demonstrate the practicality of the approach.

Table IX shows the internal flash, running memory SRAM, and complexity of translating the classical network model into code executable by embedded devices for MCU execution. It is clear that Xception requires the most resources and also has the highest model complexity relative to the other models. In contrast, the model proposed in this article requires 120 kB of ROM, 59.04 kB of SRAM, and 1.6M of Macc, respectively. Owing to the limited size of the SRAM in the embedded device, we make minor changes to the proposed model.

TABLE X
DETAILS OF THE MODEL STRUCTURE FOR HARDWARE DEPLOYMENT

Layer	Output Size
Input	1×1500
Conv1d	16×750
Max-Pooling	16×375
Group-Conv1	32×188
Group-Conv2	64×94
ECA	64×94
BN	64×94
LSTM	64×32
GAP	64×1
FC	64×1
SoftMax	2×1

Table X shows the changes in a network structure, where one group convolutional layer is removed and the layers of the LSTM are adjusted relative to the original network structure. Table VIII shows the test results, and it is worth noting that the number of samples is 200 for dataset 1 and 50 for dataset 2. It is clear that running the model on the PC is much faster than the PCG sensor, and yet, the average run time of the sensor per sample is only 1.58 s, which is acceptable for practical applications. In terms of power consumption, the sensor has around 129.9 mW for data acquisition and an average of 245.1 mW for performing the classification task. Meanwhile, the classification accuracy of the sensor achieves 89.8% and 95.6% on the two datasets, a decrease of 3.3% and 3.8%, respectively, relative to the best result of the original model.

VI. DISCUSSION

In this work, we propose a lightweight DL network model for the classification of abnormal heart sounds and design a PCG sensor device for deploying the model. We have carried out experiments in three aspects.

First, to ensure the robustness of the model, data from different groups of people (healthy and patients) are mixed together for training and testing. Two public datasets are utilized to evaluate the performance of the proposed method. We first discuss the binary classification issues. From Table V, we can see that the proposed model achieves the best results, with a 2.9%, 3.0%, and 0.3% improvement compared with the baseline in accuracy, specificity, and $F1$ score, respectively, but the sensitivity is lower than the baseline. The reason for this is that the training dataset is very imbalanced, and the test samples have a high proportion of negative samples (as obtained from Fig. 7), leading to an obvious difference in their classification accuracies of the positive samples. In addition, for the pathological performances, the baseline values are not designed due to the small samples of dataset 2 and the already high metrics of the existing algorithms. However, our proposed model remains achieving better results than in previous work, especially in terms of accuracy, sensitivity, and $F1$ score. In summary, the performance of the proposed model is more balanced in the classification of heart sound diseases. This comes, as the model automatically extracts data features from multiple channels and multiple dimensions and does not rely on manual features. Thus, the proposed model

can be applied to the classification of heart sounds in terms of abnormal and pathological classification.

Second, we propose a GCT to design lightweight end-to-end models. In general, most of the previous work has been based on feature engineering, with models of high complexity and computational power. This makes it difficult to deploy the models in some low-power, low-computing embedded devices. As can be seen from Fig. 9, the proposed model achieves better results in terms of accuracy, the number of parameters, and the number of operations compared with classical lightweight models such as MobileNet. Furthermore, Table VII shows that the proposed model achieves a more balanced performance compared with the previous state-of-the-art work, with Macc improving by 1.7% compared with the baseline, while the number of parameters and operations is only 29.36% and 45.34% of the baseline. It is worth noting that the Macc indicator remains 7% lower than in [13], because its input is 2-D features, and the model may have a better learning ability. The proposed model, however, has a great improvement in terms of the number of parameters and the number of operations. Considering the practical deployment of the algorithm, the proposed model shows excellent overall performance and is more suitable for implementation in embedded devices with low-computational power.

Third, we design the PCG sensor to deploy the proposed method, which captures the heart sound signal precisely. In the experiments, we first compare the run time of the algorithm, which has a longer running time than the PC, as shown in Table VIII. However, we consider this time acceptable in practical applications, because it achieves nearly real-time classification. Moreover, in terms of power consumption, the device consumes an average of just 66.2 mA when performing classification tasks; thus, it can work continuously for around 4.5 h when the device carries a battery capacity of 300 mA. Furthermore, it can work continuously for around 8.5 h in the single data collection scenario, a result that is extremely important in the long-term monitoring environment. While we try to reduce the complexity of the model, we have had to simplify the proposed model, since it has a large LSTM layer and convolutional layers, which cannot be deployed directly on hardware devices. Fortunately, the modified model still achieves a good performance, which shows the validity of our work. Wearable devices are prone to power consumption and complexity issues, making low-power consumption and computational power crucial for long-term usability. Furthermore, reducing the number of FLOPs translates to lower power consumption on the same hardware. Although we have not directly deployed the models from other work in the sensors designed in this article, Table VII clearly demonstrates that our proposed model has significantly lower FLOPs than previous works. This means that our proposed model achieves excellent performance while consuming lower power. Also, the fact that the model proposed in this article can run smoothly in the designed sensor also shows that the work in this article is meaningful.

Despite the promising results achieved by our proposed algorithm and its successful deployment on embedded devices, there remain some limitations to improve. The training dataset is derived from a publicly available dataset rather than collected with the sensor by our design, which can degrade the performance of the model in practice. Future work requires the creation of specialized datasets. In addition, the interpretability of the models needs to be improved. Even though comparable AI devices play an auxiliary diagnostic role for doctors, the basic principles involved are also important.

VII. Conclusion

In this work, we design low-power PCG sensors that can run the DL model. Furthermore, we design a lightweight end-to-end model that achieves excellent performance while significantly reducing its complexity. Compared with previous work, our proposed model offers the advantage of direct deployment on PCG sensors with limited processing power, which holds significant importance for future practical application scenarios. Subsequent work will concentrate on enhancing the performance of the PCG sensor.

References

[1] WHO News, Geneva, Switzerland. *Who Reveals Leading Causes of Death and Disability Worldwide: 2000–2019*. Accessed: Dec. 9, 2020. [Online]. Available: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019

[2] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9393–9400, Nov. 2018.

[3] S. Li, F. Li, S. Tang, and W. Xiong, "A review of computer-aided heart sound detection techniques," *BioMed Res. Int.*, vol. 2020, pp. 1–10, Jan. 2020.

[4] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021.

[5] L. Zhu, K. Qian, Z. Wang, B. Hu, Y. Yamamoto, and B. W. Schuller, "Heart sound classification based on residual shrinkage networks," in *Proc. EMBC*, Glasgow, U.K., Jul. 2022, pp. 4469–4472.

[6] W. Qiu et al., "A federated learning paradigm for heart sound classification," in *Proc. EMBC*, Glasgow, U.K., Jul. 2022, pp. 1045–1048.

[7] F. Noman, C.-M. Ting, S.-H. Salleh, and H. Ombao, "Short-segment heart sound classification using an ensemble of deep convolutional neural networks," in *Proc. ICASSP*, Brighton, U.K., May 2019, pp. 1318–1322.

[8] K. Qian, Z. Zhang, Y. Yamamoto, and B. W. Schuller, "Artificial intelligence Internet of Things for the elderly: From assisted living to health-care monitoring," *IEEE Signal Process. Mag.*, vol. 38, no. 4, pp. 78–88, Jul. 2021.

[9] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 8, pp. 2189–2198, Aug. 2020.

[10] T. T. K. Munia et al., "Heart sound classification from wavelet decomposed signal using morphological and statistical features," in *Proc. CinC*, Vancouver, BC, Canada, Sep. 2016, pp. 597–600.

[11] S. Li, F. Li, S. Tang, and F. Luo, "Heart sounds classification based on feature fusion using lightweight neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[12] V. Arora, R. Leekha, R. Singh, and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *Mod. Phys. Lett. B*, vol. 33, no. 26, Sep. 2019, Art. no. 1950321.

[13] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, Oct. 2020.

[14] S.-Y. Lee, P.-W. Huang, J.-R. Chiou, C. Tsou, Y.-Y. Liao, and J.-Y. Chen, "Electrocardiogram and phonocardiogram monitoring system for cardiac auscultation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1471–1482, Dec. 2019.

[15] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, Feb. 2017.

[16] Z. Dokur and T. Ölmez, "Feature determination for heart sounds based on divergence analysis," *Digit. Signal Process.*, vol. 19, no. 3, pp. 521–531, May 2009.

[17] Z. Wang, Z. Bao, K. Qian, B. Hu, B. W. Schuller, and Y. Yamamoto, "Learning optimal time-frequency representations for heart sound: A comparative study," in *Proc. CSMT*. Hangzhou, China: Springer, 2023, pp. 93–104.

[18] H. Li et al., "A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103733.

[19] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features," *IEEE Sensors Lett.*, vol. 3, no. 12, pp. 1–4, Dec. 2019.

[20] X. Huai, S. Notsu, D. Choi, P. Siriaraya, and N. Kuwahara, "Development of wearable heart sound collection device," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, pp. 1–8, 2021.

[21] Y. Ning, M. Zhang, Y. Lang, Y. Gong, X. Yang, and W. Pang, "Electronic stethoscope based on triangular cantilever piezoelectric bimorph MEMS transducers," *J. Microelectromech. Syst.*, vol. 31, no. 3, pp. 450–456, Jun. 2022.

[22] M. E. H. Chowdhury et al., "Real-time smart-digital stethoscope system for heart diseases monitoring," *Sensors*, vol. 19, no. 12, p. 2781, Jun. 2019.

[23] S. Zhang, R. Zhang, S. Chang, C. Liu, and X. Sha, "A low-noise-level heart sound system based on novel thorax-integration head design and wavelet denoising algorithm," *Micromachines*, vol. 10, no. 12, p. 885, Dec. 2019.

[24] P. Gupta, M. J. Moghimi, Y. Jeong, D. Gupta, O. T. Inan, and F. Ayazi, "Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, Feb. 2020.

[25] C. Deng, X. Fang, X. Wang, and K. Law, "Software orchestrated and hardware accelerated artificial intelligence: Toward low latency edge computing," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 110–117, Aug. 2022.

[26] H. Zairi, M. K. Talha, K. Meddah, and S. O. Slimane, "FPGA-based system for artificial neural network arrhythmia classification," *Neural Comput. Appl.*, vol. 32, no. 8, pp. 4105–4120, Apr. 2020.

[27] M. Alfaro-Ponce, I. Chairez, and R. Etienne-Cummings, "Automatic detection of electrocardiographic arrhythmias by parallel continuous neural networks implemented in FPGA," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 363–375, Feb. 2019.

[28] W.-S. Jhong et al., "Deep learning hardware/software co-design for heart sound classification," in *Proc. ISOCC*, Yeosu, South Korea, Oct. 2020, pp. 27–28.

[29] G. Li, H. Yang, T. Guo, and W. Wang, "Implementation and acceleration scheme of heart sound classification algorithm based on SOC-FPGA," in *Proc. BIC*, Harbin, China, Jan. 2022, pp. 258–266.

[30] S. A. Fattah et al., "Stetho-phone: Low-cost digital stethoscope for remote personalized healthcare," in *Proc. GHTC*, San Jose, CA, USA, Oct. 2017, pp. 1–7.

[31] F. Dong et al., "Machine listening for heart status monitoring: Introducing and benchmarking HSS—The heart sounds Shenzhen corpus," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2082–2092, Jul. 2020.

[32] B. M. Whitaker and D. V. Anderson, "Heart sound classification via sparse coding," in *Proc. CinC*, Vancouver, BC, Canada, Sep. 2016, pp. 805–808.

[33] S. A. Singh and S. Majumder, "Classification of unsegmented heart sound recording using KNN classifier," *J. Mech. Med. Biol.*, vol. 19, no. 4, Jun. 2019, Art. no. 1950025.

[34] C. C. Balili, Ma. C. C. Sobrepena, and P. C. Naval, "Classification of heart sounds using discrete and continuous wavelet transform and random forests," in *Proc. ACPR*, Kuala Lumpur, Malaysia, Nov. 2015, pp. 655–659.

[35] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Support vector machine hidden semi-Markov model-based heart sound segmentation," in *Proc. Cinc*, Cambridge, MA, USA, Sep. 2014, pp. 625–628.

[36] 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.

[37] G. Tian et al., "Imbalanced heart sound signal classification based on two-stage trained DsaNet," *Cogn. Comput.*, vol. 14, pp. 1378–1391, Mar. 2022.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[39] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, Jun. 2020.

[40] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, Nov. 2018.

[41] F. Chakir, A. Jilbab, C. Nacir, and A. Hammouch, "Phonocardiogram signals processing approach for PASCAL classifying heart sounds challenge," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1149–1155, Sep. 2018.

[42] C. Liu et al., "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Dec. 2016.

[43] G. D. Clifford et al., "Classification of normal/abnormal heart sound recordings: The PhysioNet/computing in cardiology challenge 2016," in *Proc. Comput. Cardiol. Conf. (CinC)*, Cambridge, MA, USA, Sep. 2016, pp. 609–612.

[44] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. CVPR*, Seattle, WA, USA, Jun. 2020, pp. 11531–11539.

[45] T. Liu et al., "Spatial channel attention for deep convolutional neural networks," *Mathematics*, vol. 10, no. 10, p. 1750, May 2022.

[46] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. ECCV*, Munich, Germany, 2018, pp. 116–131.

[47] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[48] F. B. Azam, M. I. Ansari, I. Mclane, and T. Hasan, "Heart sound classification considering additive noise and convolutional distortion," 2021, *arXiv:2106.01865*.

[49] A. I. Humayun, S. Ghaffarzadegan, Z. Feng, and T. Hasan, "Learning front-end filter-bank parameters using convolutional neural networks for abnormal heart sound detection," in *Proc. EMBC*, Honolulu, HI, USA, Jul. 2018, pp. 1408–1411.

[50] S. L. Oh et al., "Classification of heart sound signals using a novel deep WaveNet model," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105604.

[51] N. Baghel, M. K. Dutta, and R. Burget, "Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, Art. no. 105750.

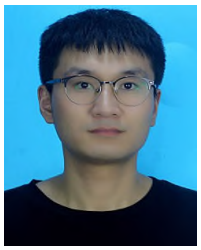
[52] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105940.

[53] 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.

[54] A. Howard et al., "Searching for MobileNetV3," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 1314–1324.

[55] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 1251–1258.

[56] 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, no. 8, pp. 1671–1684, Jul. 2017.

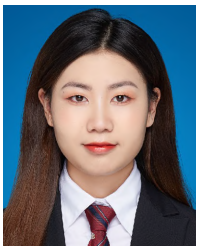


Lixian Zhu received the M.S. degree in communication and information system from the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China.

His research interests include biosensor technology, biomedical signal processing, and deep learning.

Wanyong Qiu received the M.S. degree from the School of Computer Science and Engineering, Northwest Normal University, Lanzhou, China, in 2021. He is currently pursuing the Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China.

His research interests include computer audition, federated learning, and healthcare information security.



Yu Ma received the M.S. degree in communication and information systems from the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China, in 2022. She is currently pursuing the Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China.

Her research interests include multimodal fusion, computational psychophysiology, and deep learning.

Fuze Tian received the Ph.D. degree in technology of computer application from the School of Information Science and Engineering, Lanzhou University, Lanzhou, China, in 2022.

He is currently a Post-Doctoral Researcher at the Brain Health Engineering Laboratory, Institute of Engineering Medicine, Beijing Institute of Technology, Beijing, China. His research fields include brain–computer interfaces, biosensor technology, and biomedical signal processing.

Mengkai Sun received the M.S. degree in data science from the Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane, QLD, Australia, in 2021. He is currently pursuing the Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China.

His research interests include computer audition and meta learning.

Zhihua Wang received the bachelor's degree from the China University of Mining and Technology (CUMT), Beijing, China, in 2018.

He has been taking successively a master–doctoral program of study for the doctoral degree at the School of Mechatronic Engineering, CUMT, since 2018. Since 2021, he has been funded by the China Scholarship Council (CSC) as a Special Research Student at the Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Tokyo, Japan. His research interests are computer audition, interpretable artificial intelligence (AI), acoustic signal analysis and processing, pattern recognition, and machine learning.

Kun Qian (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and information technology from Technische Universität München (TUM), Munich, Germany, in 2018, with a focus on automatic general audio signal classification.

Since 2021, he has been appointed as a (Full) Professor at the Beijing Institute of Technology, Beijing, China. He has authored or coauthored more than 90 publications in peer-reviewed journals and conference proceedings having received more than 1.6k citations (H-index 22).

Dr. Qian has a strong collaboration connection to prestigious universities in Germany, the United Kingdom, Japan, Singapore, and the United States. He serves as an Associate Editor for IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *Frontiers in Digital Health*, and *BIO Integration*.

Bin Hu (Fellow, IEEE) received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 1998.

He is a Full Professor and the Dean of the School of Medical Technology at Beijing Institute of Technology, China. He is a National Distinguished Expert, the Chief Scientist of 973 as well as the National Advanced Worker in 2020. He is the Principal Investigator for big grants such as the National Transformative Technology "Early Recognition and Intervention Technology of Mental Disorders Based on Psychophysiological Multimodal Information," which have greatly promoted the development of objective, quantitative diagnosis and non-drug interventions for mental disorders. He (co-)authored more than 400 publications in peer reviewed books, journals, and conference proceedings leading to more than 12 k citations (h-index 58).

Dr. Hu is the Fellow of IET. He is a Member of the Steering Council of the ACM China Council and the Vice-Chair of the China Committee of the International Society for Social Neuroscience. He serves as the Editor-in-Chief for IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS and an Associate Editor for IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. He is a member of the Computer Science Teaching and Steering Committee as well as the Science and Technology Committee. He is a recipient of many research awards, including the 2014 China Overseas Innovation Talent Award, the 2016 Chinese Ministry of Education Technology Invention Award, the 2018 Chinese National Technology Invention Award, and the 2019 WIPO-CNIPA Award for Chinese Outstanding Patented Invention. He is also the TC Co-Chair of computational psychophysiology in the IEEE Systems, Man, and Cybernetics Society (SMC), and the TC Co-Chair of cognitive computing in IEEE SMC.

Yoshiharu Yamamoto (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in education from The University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1990, respectively.

Since 2000, he has been a Professor at the Graduate School of Education, The University of Tokyo, where he is teaching and researching physiological bases of health sciences and education. He has authored or coauthored more than 230 publications in peer-reviewed books, journals, and conference proceedings leading to more than 12k citations (H-index 59). His research interests include biomedical signal processing, nonlinear and statistical biodynamics, and health informatics.

Dr. Yamamoto is an Editorial Board Member of *Technology* and *Biomedical Physics and Engineering Express*. He is the President of the Healthcare IoT Consortium, Japan. He is an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.

Björn W. Schuller (Fellow, IEEE) received the Diploma, Ph.D., and Habilitation degrees in electrical engineering and information technology from Technische Universität München (TUM), Munich, Germany, in 1999, 2006, and 2012, respectively, with a focus on signal processing and machine intelligence.

He was an Adjunct Teaching Professor in the subject area of signal processing and machine intelligence in electrical engineering and information technology with TUM. He is a tenured Full Professor heading the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany, and a Professor of artificial intelligence (AI) heading the Group on Language, Audio, and Music (GLAM), Department of Computing, Imperial College London, London, U.K. He has authored or coauthored five books and more than 1000 publications in peer-reviewed books, journals, and conference proceedings leading to more than 49k citations (H-index 101).

Dr. Schuller is a Senior Member of the ACM. He is a fellow of the International Speech Communication Association (ISCA) and the British Computer Society (BCS). He is the Golden Core Awardee of the IEEE Computer Society. He is the President Emeritus of the Advancement of Affective Computing (AAAC). He is a former Editor-in-Chief of IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. He is a Field Chief Editor of *Frontiers in Digital Health*.