A Systematic Review of Heart Sound Detection Algorithms: Experimental Results and Insights

Xiuli Bi, Shizhan Tang, Bin Xiao, Weisheng Li, *Member, IEEE*, Xinbo Gao, and Pietro Liò

***Abstract*— Automatic heart sound detection plays a vital role in the early detection of cardiovascular diseases (CVDs). In recent years, many heart sound detection algorithms have been proposed, becoming a popular research topic in medical diagnosis. In this article, our goal is to answer which heart sound detection algorithms and heart sound features perform best in which situations and to explore the problems in existing research. This work will provide ideas for subsequent research in this area. We achieve this aim through a standard process that includes preprocessing, segmentation, feature extraction, and classification. We discuss the length and overlap rate in segmentation and analyze the classification methods, especially the nine most salient heart sound features. The performance of the existing techniques is evaluated in different datasets and at the same level of comparison. The experiments show that the segmentation length includes at least one complete period, and we should use a large overlap rate in the segmentation phase. For feature extraction, time-domain feature (TIME) and fast Fourier transform (FFT) features based on single independent variables perform better in deep models and traditional classifiers. Short-time Fourier transform (STFT), continuous wavelet transform (CWT), and S-transform (ST) features based on double independent variables can perform well in all classification models; Mel spectrum/Mel frequency cepstrum coefficient (MFCC) is only better on deep neural networks (DNNs). The best performance is achieved by combining TIME or MFCC with DNNs.**

***Index Terms*— Classification model, comparative study, feature extraction, heart sound detection.**

I. Introduction

CARDIOVASCULAR diseases (CVDs) are diseases with high morbidity and mortality worldwide. According to statistics, 20.5 million people died from CVD in 2021, accounting for about one-third of all global deaths [1].

Received 26 June 2024; revised 2 September 2024; accepted 15 September 2024. Date of publication 3 March 2025; date of current version 19 March 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62221005 and Grant 62172067; in part by the National Major Scientific Research Instrument Development Project of China under Grant 62027827; and in part by the Natural Science Foundation of Chongqing for Distinguished Young Scholars under Grant CSTB2022NSCQ-JQX0001. The Associate Editor coordinating the review process was Dr. Lin Xu. *(Corresponding author: Bin Xiao.)*

Xiuli Bi, Shizhan Tang, Weisheng Li, and Xinbo Gao are with the Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: bixl@cqupt.edu.cn; s210231165@stu.cqupt.edu.cn; liws@cqupt.edu.cn; gaoxb@cqupt.edu.cn).

Bin Xiao is with the Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with Jinan Inspur Data Technology Company Ltd., Jinan 250101, China (e-mail: xiaobin@cqupt.edu.cn).

Pietro Liò is with the Department of Computer Science and Technology, University of Cambridge, CB3 0FD Cambridge, U.K. (e-mail: pl219@cam.ac.uk).

Digital Object Identifier 10.1109/TIM.2025.3547082

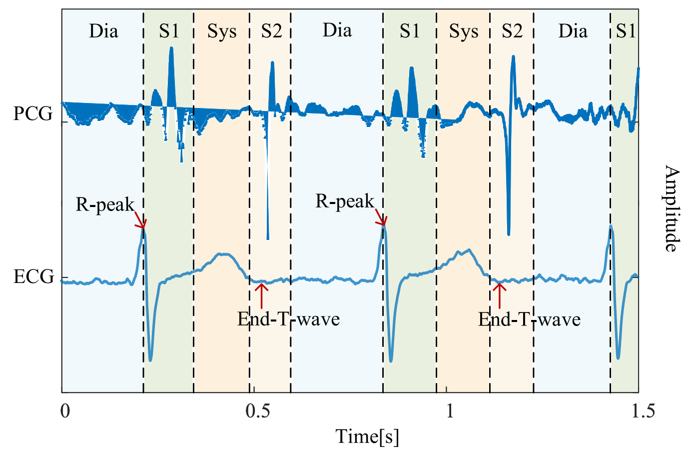

Fig. 1. PCG (dark color) with simultaneous ECG (light color) recording and the four states of the PCG recording: S1, the systole (Sys), and S2, the diastole (Dia).

Detecting and diagnosing CVDs early can significantly reduce mortality rates [2]. Using a stethoscope on the chest to capture heart sounds is an old, noninvasive, and cost-effective method for the early detection of various CVDs. However, heart sound auscultation relies heavily on the expertise of physicians, which is developed through extensive training and clinical experience. Despite this, there are instances where even trained physicians may misdiagnose [3]. This issue is particularly pronounced in regions with limited healthcare resources, such as low- and middle-income areas. Therefore, developing a computer-based heart sound analysis tool holds immense promise in aiding physicians and patients in diagnosing heart diseases more accurately and efficiently. Such a tool could complement healthcare providers' skills, especially in resource-constrained settings, improving the effectiveness of early diagnosis strategies.

The exploration of heart sounds dates back to as early as 1963 [4]. People record heart states using electronic devices and then plot over a time axis to create a phonocardiogram (PCG) or an electrocardiogram (ECG) as information for assessing various CVDs. As depicted in Fig. 1, the typical PCG waveform encompasses four distinct states: the first heart sound (S1), the second heart sound (S2), systole, and diastole. S1 occurs during systole, signaling the initiation of ventricular contraction, while S2 occurs during diastole, marking the onset of ventricular relaxation. If there appears a third heart sound (S3) and a fourth heart sound (S4) in the PCG, which are extra gallop sounds in unhealthy adult hearts and are essential clues for assessing various CVDs. Although ECG can also

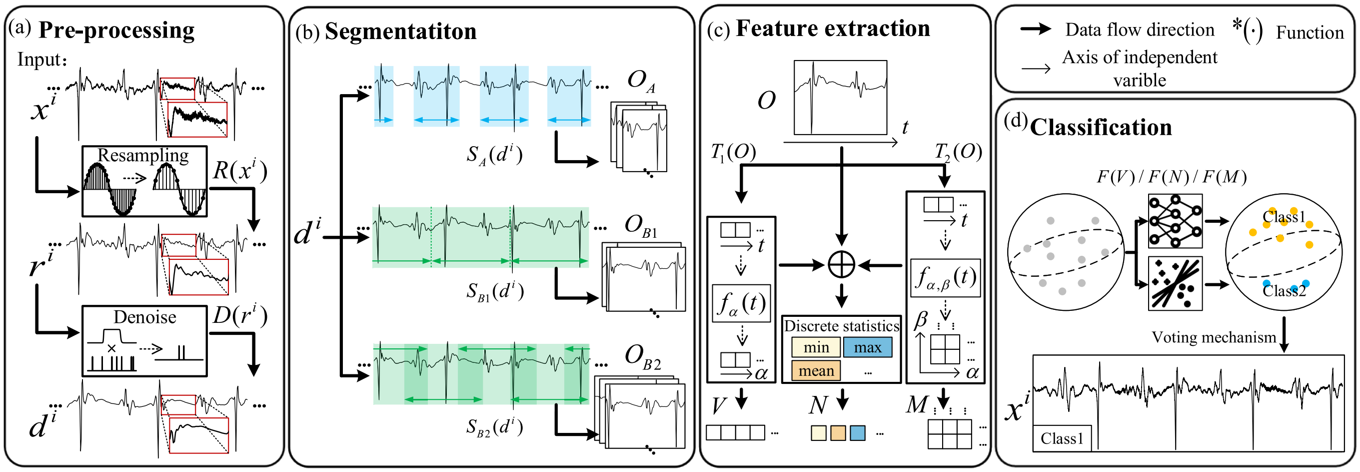

Fig. 2. There are four standard steps for heart sound detection. (a) Resampling and denoising operation in preprocessing and their functions. For example, they make signal sampling frequency reduction and high-frequency noise disappearance, as illustrated in the red line box. (b) Three segmentation cases that are different in the number and size of slices in each case. (c) Two feature extraction approaches for extracting heart sound slices. (d) Classification, which classifies the slices by features and votes the type of the original heart sound. Each standard step has some methodological choices, which will be described in detail in Section II.

diagnose CVDs, PCG is more effective than ECG in showing congenital heart defects, such as heart valve defects [5], [6].

Over the decades, with significant advancements in science and technology, particularly with the introduction of PhysioNet/CinC Challenge 2016 [7], there has been a notable rise in approaches for heart sound detection in various scholarly publications. Currently, the reviews that exist on heart sound detection do not summarize the field well. Chen et al. [8] only summarize the application of neural networks in heart sound detection. Dwivedi et al. [9] only use tables to show the previous algorithms and briefly discuss whether these methods suit wearable devices. While Chen et al. [10] extensively review many algorithms, they omit some experiments for summarization validation. Roy et al. [11] enumerate some existing heart sound algorithms and investigate the classification performance of convolution-based classification models.

This article aims to review and summarize various classical heart sound detection algorithms, reproduce them, and offer valuable perspectives for future research. We have collected heart sound segmentation and detection publications from ScienceDirect, IEEEXplore, SpringerLink, ResearchGate, and other databases. We select over 100 articles that are promising methods for heart sound detection using different feature extractions, datasets, and evaluation metrics. Since we consider a competitive algorithm capable of coping with all scenarios and realistic scenarios are more complex, we define a standard pipeline for the heart sound detection methods using distinctive, challenging public datasets and uniform metrics. In addition, based on the above settings, we investigate the variations arising from differences in length and overlap rate in heart sound segmentation, an aspect often overlooked in many articles. We also implement the nine most representative features from over 100 articles and discuss their performance on heart sound detection with different classification models.

We organize this article as follows. In Section II, we present a discussion of existing advanced heart sound segmentation and classification approaches in a standard process. Section III presents the existing evaluation metrics for heart sound detection and the available public databases. The experiments and analysis are given in Section IV. We systematically summarize the observed experimental results in Section V. Section VI contains a summary and closing remarks.

II. Standard Workflow for Heart Sound Classification

Since a standard heart sound detection process will help us discuss the existing works more clearly, we define a standard process, as shown in Fig. 2. Since most existing approaches follow the same process: preprocessing, feature extraction, and classification, while heart sound segmentation is divided as a separate study domain or in preprocessing for serving the downstream task of heart sound better, we consider it as an individual step between preprocessing and feature extraction in this article.

In this standard heart sound detection process, the original heart sounds from the human body will first be preprocessed, which includes resampling the heart sound signal while reducing the noise. Following this is the heart sound segmentation, involving exact segmentation using algorithms or direct equal-length segmentation using sliding windows. It is worth noting that exact segmentation may also involve specialized feature extraction approaches to enhance segmentation accuracy. Then, features based on single or double independent variables are extracted from the slices, i.e., the heart sounds are represented in different features. Finally, each heart sound representation after feature extraction is classified using various classification models. A voting mechanism is then applied to determine whether a whole heart sound is normal. If the number of abnormal slices across the whole heart sound is more significant than the number of normal slices, it is judged abnormal, and vice versa. We will discuss detailed descriptions and methodological examples of each step in Sections II-A–II-D to provide a comprehensive illustration for subsequent workflow-based experiments.

A. Preprocessing

1) Resample: Humans' average heart sound signal frequency is about 800 Hz, and heart sound collectors often set a high sampling frequency to collect high-quality heart sound data. For instance, the sampling frequency of our infant dataset is 44.1 kHz. However, when we use the heart sound dataset, the heart sound data are resampled to reduce the computers' load and make the algorithm feasible. The resampling frequency in most works is typically set to 2000 [12], [13], [14], [15], [16], [17] or 1000 Hz [18], [19], [20], [21]. Since 2000 Hz is larger than twice the human's average heart sound signal frequency, satisfying the Nyquist theorem, it can losslessly keep the original heart sound signals.

2) Denoise: Since the collection environment cannot be entirely noise-free for heart sounds, such as external natural noises, sounds of other organs vibrating in the human body, and sounds of machines, removing these noises in the preprocessing stage can improve the performance of heart sound detection [22]. These noises mainly exist in the high-frequency and some low-frequency regions of heart sound, while the features of heart sounds are primarily concentrated in the low- and middle-frequency regions. Many works directly use filterings or denoising algorithms to filter the heart sounds for noise reduction, for instance, Butterworth bandpass filter [23], [24], [25], [26], Chebyshev filter [27], wavelet threshold denoising algorithms [28], [29], and Wiener spectral subtraction [22].

B. Segmentation

One heart sound period has four states: S1, systole, S2, and diastole. Since most of the pathological features can be extracted in S1 and/or S2, locating the S1 and S2 regions is crucial. Heart sound segmentation is proposed to segment the S1 and S2 regions. It generally uses the ECG signals as a standard. As shown in Fig. 1, P-peaks of the ECG correspond to the beginning of S1, and the center of S2 corresponds to the end of T-peaks. Meanwhile, heart sound segmentation constitutes a crucial step within heart sound detection algorithms. If a heart sound detection algorithm is based on precise segmentation, superior classification outcomes are primarily contingent upon the accuracy of the segmentation results.

Before deep learning was used for heart sound detection, traditional heart sound detection approaches used exact segmentation algorithms [30], [31], and these exact segmentation algorithms are almost based on feature extraction. Milani et al. [32] classify these feature-based exact segmentation approaches into four cases: wavelet transform (WT) [18], fractal decomposition [33], Hilbert envelope algorithm [26], [34], and Shannon energy envelope [35]. With deep learning developments, due to the powerful learning ability of deep neural networks (DNNs), we can achieve a better classification by only equal-length segmentation of heart sounds [36], [37], [38]. Although exact segmentation is gradually becoming obsolete in heart sound detection, precisely segmenting the four states of one heart sound period is still of practical importance for analyzing heart sounds.

TABLE I
DIFFERENT TYPES OF FEATURES APPLIED IN HEART SOUND DETECTION

Type	Feature	Method
SIV-based feature	TIME	[40], [12], [41], [42], [37], [43], [44], [38], [45], [46], [47], [48], [49], [50], [51] [52], [53]
	DFT	[54], [55], [56]
	PSD	[57], [55]
DIV-based feature	STFT	[16], [23], [24], [58], [59], [60], [13], [61], [62], [63] [64], [65], [66]
	Mel/MFCC	[67], [68], [69], [70], [71], [72], [73], [74], [60], [75], [76], [77] [78]
	WT	[28], [79], [80], [81], [25], [82], [83], [84]
	ST	[85], [86], [27]
	WPD	[87], [88], [14], [89]
Multi-feature	-	[90], [91], [92], [20], [21], [93], [94], [95], [19], [96], [97], [98]
Others	-	[36], [99], [100], [101], [102], [17], [103], [104]

C. Feature Extraction

Feature extraction is essential for heart sound detection [39]. We list the various representative approaches in Table I, according to different features used for heart sound detection. SIV and DIV are abbreviations for single and double independent variables, respectively. We will mainly introduce these two feature methods, briefly describing the multifeature method and other specialized approaches. Finally, we will provide a concise summary.

1) SIV-Based Feature: The SIV-based feature is a 1-D vector whose values change as a single variable change. We investigate three SIV-based features commonly used for heart sound detection, almost representing the entire field.

a) Time-domain feature: When the amplitude of PCGs transforms with time, PCGs are a direct form of heart sound representation. The time-domain feature (TIME) often exists as a discrete statistical feature. Researchers extract the peaks, means, standard deviations, periods, and identical state spacing of PCGs as features of heart sounds or apply some transformation approaches to generate feature vectors to express PCGs. For instance, Badlaoui and Hammouch [40] extract the standard deviation of the PCG to distinguish between positive and abnormal heart sounds. Khaled et al. [91] extract a vector feature with a length of 27 from PCGs, where three terms are the mean, median, and standard deviation of PCGs on the time domain. Chakir et al. [92] exploit discrete statistical features such as period, energy, and entropy between S1 and S2 of PCGs. Similarly, Zannat et al. [105] suggest that the period interval time between (S1, S2) and the region (A1, A2) can reflect a wide range of pathologies in the heart sound signal, where the A1 and A2 are heart sounds in the aortic valve region. After deep learning is applied to heart sound detection, [12], [37], [38], [41], [42], [43], [44], [45], [46], [47], [49], [50], [51], and [52] can use PCGs as inputs directly. These works greatly improve the performance of heart sound detection. Moreover, Li et al. [48] incorporate the features of PCGs and the features of corresponding ECGs.

b) Discrete Fourier transformation: Discrete Fourier transform (DFT) is the most classical approach in signal processing. Fast Fourier transform (FFT) is an algorithm that speeds up DFT calculations. DFT transforms signals from the time domain to the frequency domain, which can study the frequency components present in signals. Since using DFT alone has limitations for processing nonsmooth signals, and PCG is also a nonsmooth signal, few methods use DFT alone. DFT is usually used as part of a multifeature or as an intermediate step. Saraçoğlu [54] and Uğuz [55] utilize the DFT feature extraction and the principal component analysis (PCA) method for heart sound detection. Yadav et al. [56] directly use DFT transform PCGs to the frequency domain and then calculate the mean and maximum values ratio in the frequency domain to detect abnormal heart sounds.

c) Power spectral density: Power spectral density (PSD) is a physical quantity that characterizes the power energy of a signal as a function of frequency and is often used to study vibration signals. Saputra et al. [57] use autoregressive PSD (AR-PSD) to transform the original PCG into the frequency domain and then extract features from the frequency domain. In our search, PSD features are not often used in heart sound detection alone but often in combination with other features, such as [90].

2) DIV-Based Feature: Unlike SIV-based features, DIV-based features can contain more information about the PCG. Mapping heart sounds to the time–frequency domain is the dominant idea behind these features. Soeta and Bito [106] illustrate the effectiveness of time–frequency analysis in disease analysis. Here, we summarize five DIV-based features that almost contain the features used in the existing heart sound detection.

a) Short-time Fourier transform: The short-time Fourier transform (STFT) maps the signal to the time–frequency domain and is one of the classical approaches in signal research in the time–frequency domain. STFT solves the problem of the Fourier transform not being able to characterize nonstationary signals well. We can analyze the variation of the signal-frequency components over time from the spectrograms obtained by STFT and obtain information in both frequency and time domains. STFT is widely used in heart sound detection studies. For example, Zhang et al. [16], [23] use STFT to extract the features of the heart sound signal and perform the heart sound detection by different dimensionality reduction approaches: partial least-squares regression (PLSR) and tensor decomposition (TD). Markaki et al. [24] redistribute the features generated by STFT to overcome the resolution tradeoff between time and frequency in the STFT method. In [58], [59], [60], [63], [64], [65], and [107], all of them directly use the spectral map generated by STFT as the features of the input classifier. Her and Chiu [13] use a method similar to STFT, dividing the region and then using the FFT to extract the time–frequency features. Nilanon et al. [61] use PSD instead of DFT in STFT as a characteristic of heart sounds. Ismail et al. [62] and Maity et al. [66] use a transfer learning approach to learn feature fine-tuning networks from the spectrograms generated by STFT.

b) Mel spectrum/Mel frequency cepstrum coefficient: In speech recognition and speaker recognition, the most commonly used features are the Mel spectrum and Mel frequency cepstrum coefficients (MFCCs). The feature extraction method uses the study of the human ear hearing mechanism, which is more in line with the hearing characteristics of the human ear and has better recognition characteristics for low signal-to-noise ratio signals. Similarly, this feature extraction method is also heavily used in heart sound detection. González Ortiz et al. [67] combines MFCC and dynamic time warping to improve the classifier's performance. Quiceno-Manrique et al. [68] compare the performance of several matrix-based features and conclude that MFCC works better than several other feature extraction approaches. Hamidi et al. [69] fuse the MFCC features and the fractal dimension of the PCG signal as features. References [60], [70], [71], [72], [73], [74], and [108] directly use MFCC as the feature input to the classifier. Also, in [75] and [76], the Mel spectrum is used directly as a feature. Barnawi et al. [77] use the transfer learning method to fine-tune the pretrained VGG19 network. Abbas et al. [78] explore the utilization of spectrograms and MFCCs for signal processing and visual description.

c) Wavelet transform: WT is an improvement of STFT. Since the window size of STFT is fixed, we cannot consider a small window for high frequencies and a large window for low frequencies. The WT changes the basis of the Fourier transform from an infinitely long trigonometric basis to a finite-length and decaying wavelet basis (i.e., the mother wavelet). Then, the window size transformation is realized by stretching and translating the mother wavelet. de Vos and Blanckenberg [28] and Ari et al. [79] use WT to extract features to diagnose successful heart sounds in children. Uğuz [80] takes discrete WT-fused subband entropy as input patterns of the adaptive neuro-fuzzy inference system (ANFIS) classifiers; this method is effective for detecting heart valve diseases. Patidar et al. [81] propose the tunable Q-WT feature extraction method, which combines subbands obtained from tunable-Q wavelet transform (TQWT) decomposition to extract diagnostic features. The correlation between the subbands can represent the clutter of each type of PCG. Deng and Han [25] use discrete wavelets (DWTs) combined with subband entropy as classification features. Meintjes et al. [82] utilize the scale maps generated by continuous WT (CWT) to complete the S1 and S2 classifications, showing the great potential of CWT and CNN in heart sound analysis. Dhar et al. [83] directly apply the cross-wavelet generated spectral images as features. Aljohani et al. [84] use continuous wavelet and DWT as input combined with convolutional neural network approaches to predict heart sounds.

d) S-transform: The S-transform (ST), a fusion of STFT and WT, addresses the limitation of the STFT's fixed window resolution and incorporates the multiresolution analysis capability of WT while retaining connection with the Fourier spectrum. Livanos et al. [85] explore three feature extraction methods: STFT, WT, and ST, and conclude that the classification based on ST is better than both STFT and WT. Moukadem et al. [86] use the Shannon entropy of the S-transformed spectral map as a feature of PCG for classification. Hadi et al. [27] directly extract the S-transformed spectral map's maximum, minimum, mean, and standard deviation as features and input them into the ANN classifier.

e) Wavelet packet decomposition: Wavelet packet decomposition (WPD) can offer a more refined signal analysis. It addresses the issue of decreasing frequency resolution with increasing frequency inherent in wavelet decomposition by decomposing both low and high frequencies. Additionally, it introduces the concept of optimal basis function selection, enabling adaptive selection of the best basis function based on the characteristics of the analyzed signal. Safara et al. [87] propose a multilevel basis function selection method based on WPD by using different frequency ranges, noise frequencies, and energy thresholds as criteria to remove the less informative basis and retain the most informative basis in the WPD tree. Qian et al. [88], [109] use fused wavelet decomposition energy and WPD energy to classify snoring and heart sounds, respectively. WPD and statistical parameters are used by Zheng et al. [14] and Mehr et al. [89] to accomplish the heart sound detection task.

3) Multifeature: Multifeatures are derived from a combination of single independent variable-based feature or double independent variable-based features mentioned earlier. Typically, this class of methods employs traditional machine learning classifiers for classification, requiring further filtering of multiple features using feature selection approaches. We list such methods briefly in Table I. This approach typically involves utilizing dozens or even hundreds of features, necessitating professional expertise or feature selection algorithms for further refinement. This approach prevailed before the emergence of DNNs.

4) Others: Besides these methods above, there are some unique methods. Karhade et al. [100] evaluate the time–frequency-domain representations of PCG signals using both time-domain polynomial chirplet transform (TDPCT) and frequency-domain polynomial chirplet transform (FDPCT). Imani [101] uses four feature extraction methods: clustering-based feature extraction (CBFE), feature extraction using attraction point (FEUAP), feature space discriminant analysis (FSDA), and double discriminant embedding (DDE). Humayun et al. [102] directly use a DNN model to learn adaptive parametric filters. Xiao et al. [36] and Er [99] use 1-D-local binary pattern (LBP) and 1-D-local ternary pattern (LTP) as features. Yazdani et al. [17] innovate including a spectral purity index in the feature set. Whitaker and Anderson [103] transform the signal into a sparse matrix to achieve heart sound detection.

TABLE II
DIFFERENT TYPES OF CLASSIFIERS USED IN HEART SOUND DETECTION METHODS

Type	Classe	Method
TML	DT/RF	[61], [110], [21], [19]
	KNN	[40], [69], [68], [92], [101]
	SVM	[103], [24], [29], [81], [14], [40], [56], [16], [58], [59], [67], [75], [79], [25], [89], [90], [94], [111], [62]
	GMM	[112]
	DHMM	[54]
	ANN	[98], [55], [57], [57], [55], [13], [73], [28], [80], [27], [90], [91], [93], [95], [97]
DNN	CNN/RNN	[12], [41], [42], [37], [43], [44], [38], [45], [46], [47], [58], [59], [60], [61], [70], [71], [72], [74], [76], [82] [83], [88], [96], [36], [100], [102], [104], [50], [51], [52], [113], [63], [84], [64], [65] [77], [66], [53]

Dominguez-Morales et al. [104] propose a tool based on convolutional neural networks for classifying heart sounds using auditory sensors' field-programmable gate array (FPGA).

D. Classification

The classifiers used in heart sound detection tasks fall into two categories. One is the traditional machine learning (TML) classifiers used in the earliest days, such as decision tree (DT), random forest (RF), K-nearest neighbor (KNN), Gaussian mixture model (GMM), support vector machine (SVM), hidden Markov model (DHMM), and artificial neural network (ANN). As listed in Table II, it is not difficult to notice the high usage of SVM and ANN. These classifiers have shown excellent performance in the early field of heart sound detection. Another is the DNN. As listed in Table II, since the superior performance of DNN over traditional machine learning classifiers in diverse pattern recognition domains [114], [115], [116], [117], [118], coupled with the availability of large, well-labeled datasets of heart sounds [7], [43], [119], [120], [121], there has been a surge in the utilization of various deep learning frameworks for heart sound detection tasks. Notably, DNNs encompass two distinct steps: feature extraction and classification. The enhancement in the performance of DNNs in heart sound detection tasks stems from their capability to learn intricate representations. Consequently, learning PCGs using the designed model structure is the key to heart sound detection.

III. DATASETS AND EVALUATION METRICS

A. Public Dataset

Suitable datasets are the critical foundation of heart sound detection research. Over the decades, the database of heart sounds has become very rich due to the dedicated collection and selfless contributions of various organizations. We summarize the information about publicly available datasets

TABLE III
HEART SOUND DETECTION DATASETS. HERE, "SF" IS THE SAMPLING FREQUENCY, "N" IS THE NO BALANCE, AND "NR" IS THE NO REPORT

Dataset	# of file	# of Classes	SF(Hz)	Balance	Time(s)	Collection Method/Device	States
(WHSM) [121]	16	Multiple	NR	NR	9	Stethoscope	S1\Sys\S2\Dia\S3\S4
(MHSML) [120]	23	Multiple	44100	N	7-9	Not Reported(NR)	S1\Sys\S2\Dia
(CAHM) [122]	64	Multiple	NR	NR	NR	NR	NR
(PHSCCD) [119]	176(656)	4(3)	Below 195	N	1-30	iStethoscope Pro iPhone app	S1\Sys\S2\Dia
(PHSD) [43]	528	Multiple	44100	N	3-249	Digital stethoscope	S1\Sys\S2\Dia\S3\S4
(PCCD) [7]	3240	2	2000	N	5-120	Multiple database fusion	S1\Sys\S2\Dia

frequently used in the literature in Table III. The heart sound datasets used in our experiments are from PhysioNet/CinC Challenge Dataset (PCCD), Pediatric Heart Sound Dataset (PHSD), and PASCAL Heart Sound Classification Challenge Dataset (PHSCCD). First, PCCD has a large amount of complete data and is highly used. Therefore, it serves as a preferred dataset for our experiments. Furthermore, children's heart sounds differ from adult heart sounds in that they can have significant S3 status, which impacts heart sound detection, so it is necessary to explore the performance of these classification approaches on the PHSD. Finally, we use a third heart sound detection dataset (i.e., PHSCCD) to make the experiments more generalizable and credible. PHSCCD allows us to explore the performance metrics of various feature extraction approaches and classification algorithms on multiple classifications simultaneously. We select two types of subdata (normal and murmur) in the dataset as a small sample dataset. In the following, we detail the three datasets used for the experiments.

1) PHSCCD: PHSCCD consists of two subdatabases, A and B. As shown in Table III, database A is a quadruple classification dataset with 176 heart sound files, and database B is a triple classification dataset with 656 heart sound files. The length of the heart sounds ranged from 1 to 30 s. The original contest website gives a detailed description of each category in the heart sound dataset.

2) PHSD: PHSD contains 528 pediatric cardiac recordings (totaling nearly 4 h), each ranging from 3 to 249 s in duration. All PCGs in the dataset were from children aged one month to 12 years and were performed in a natural clinical setting using a ThinklabsOne digital stethoscope with a sampling frequency of 44.1 kHz and 16 bits per sample. The labels classify all patients into normal, atrial septal defect (ASD), ventricular septal defect (VSD), atrial septal defect (ASD) and ventricular septal defect (VSD), tetralogy of Fallot (TOF), and other heart-related diseases such as mitral regurgitation, atrial stenosis, and pulmonary stenosis.

3) PCCD: The PhysioNet/CinC challenge dataset is a collection of PCG records collected from seven different research groups, including an open training set and a hidden test set. The training set consists of six databases (a–f) from the other study groups. It includes 3240 heart sound recordings collected from healthy subjects and patients with various conditions, with each PCG ranging from 5 to 120 s in duration. These recordings are collected in clinical and nonclinical settings (e.g., home visits) using electronic devices. All recordings are resampled from 44.1 kHz to 2000 Hz, each containing only one PCG label. The distribution of the dataset is realistic, with a surplus of normal heart sounds over abnormal heart sounds, 2575 normal heart sounds, and 665 abnormal heart sounds.

B. Evaluation Metric

Heart sound detection involves determining whether a heart sound is normal or abnormal. The most commonly used assessment metrics in the field of heart sound detection are Accuracy (Acc), Specificity (Spe), Sensitivity (Sen), Precision (Pre), and $F1$-score ($F1$). They are defined as

$$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{1}$$

$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{2}$$

$$\mathrm{Sen} = \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{3}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{4}$$

$$F1\text{-score} = 2 \times \frac{\mathrm{Pre}\times\mathrm{Sen}}{\mathrm{Pre}+\mathrm{Sen}}. \tag{5}$$

TP is the number of true positive results, TN is the number of true negative results, FP is the number of false positive results, and FN is the number of false negative results. We consider the correct diagnosis of abnormal heart sounds recorded as true-positive samples in this article. The heart sound dataset is generally a nonequilibrium dataset. Therefore, Acc is not a suitable evaluation metric, and $F1$-score is a critical evaluation metric.

IV. EXPERIMENT AND RESULT

In this section, we perform experiments following the detailed standard process outlined in Section III-A. We propose methods for configuring two crucial factors for equal-length segmentation in DNNs: segmentation length and overlap rate. Additionally, we compare the performance benefits of SIV- and DIV-based features across various classifiers to summarize effective feature selection strategies in heart sound detection studies. Finally, we reproduce the representative heart sound detection methods to facilitate a uniform and equitable comparison of the state-of-the-art algorithms.

A. Detailed Standard Process

In Section II, we establish a standard process to provide a clear overview of the heart sound detection method. To ensure fairness in subsequent experiments, we will conduct experiments based on this standard process. Therefore, the greater details of our established standard process, which can also serve as a textual description of Fig. 2, are as follows.

1) Given a PCG recording defined as x^i, $x^i \in X$, with sampling time t^i seconds and sampling frequency F Hz, we can calculate its length $l^i = t^i \times F$ (time by sampling frequency multiplied). $X \in \mathbb{R}^{N\times L}$ is a database of heart sounds collected through an electronic stethoscope. N is the total number of heart sounds. L is the length of the heart sound in the database, but L is not the same for each heart sound, depending on the length of the sampling time.
2) For a resampling operation $R(\cdot)$, resampling the high-frequency-sampled heart sound signal to 2000 Hz [12] satisfies Nyquist theorem, such that $r^i = R(x^i)$ and $l_r^i = 2000 \times l^i/F$, where r^i is the signal after x^i has been resampled and l_r^i is the length of r^i. (Exception: When using logistic regression-HSMM-based heart sound segmentation (LR-HSMM) [18] to accurately segment the heart sounds, the original requirement is satisfied, and the signal needs to be resampled to 1000 Hz.) Afterward, a denoising method, $D(\cdot)$, is used to denoise the heart sounds, such that $d^i = D(r^i)$, where d^i is the signal after r^i has been denoised. Here, it is a sixth-order 20–950-Hz Butterworth bandpass filter [59].
3) There exist two types of segmentation method $S_A(\cdot)$ and $S_B(\cdot)$, where $S_A(\cdot)$ is the accurate segmentation method [18] and $S_B(\cdot)$ is the equal-length segmentation method. In addition, the equal-length segmentation can be divided into two cases: overlapping and nonoverlapping windows. We denote them by $S_{B1}(\cdot)$ and $S_{B2}(\cdot)$, respectively. If the length of the equal-length split window is l_{seg}, in the nonoverlapping case, for example, $S_{B1}(d^i) = O_{B1} = \{o_{B1}^1, \dots, o_{B1}^j, \dots, o_{B1}^J\}$, where o_{B1}^j is a slice, $j \in \{1, \dots, J\}$, J is the number of slices, and O_{B1} is the set of slices that split d^i by S_{B1}, $J = \lfloor l_r^i/l_{\text{seg}}\rfloor + 1$. Similarly, $S_{B2}(d^i) = O_{B2} = \{o_{B2}^1, \dots, o_{B2}^k, \dots, o_{B2}^K\}$, $K = \lfloor (l_r^i - l_{\text{seg}})/l_{\text{step}}\rfloor + 1$, where K is the number of slices and l_{step} is the split window step length, $k \in \{1, \dots, K\}$, and $S_A(d^i) = O_A = \{o_A^1, \dots, o_A^q, \dots, o_A^Q\}$, where Q is the number of periods, $q \in \{1, \dots, Q\}$. Generally, $J \leq K$. We use the output of S_{B1} as input for subsequent operations.
4) For SIV-based feature extraction approaches, we turn each slice set into a vector representation by a feature extraction algorithm $T_1(\cdot)$, such that $v^j = T_1(o_{B1}^j)$, This implies that a slice goes through an SIV-based feature extraction algorithm to get a vector representation, and $v^j \in \mathbb{R}^{1\times l_j}$ and $V = T_1(O_{B1})$, where l_j is the length of v^j and V is the set of vector representations. For DIV-based feature extraction methods, we will use this type of feature extraction algorithm $T_2(\cdot)$ to let the heart sound slices represented as a matrix spectrogram, such that $m^j = T_2(o_{B1}^j)$, $m^j \in \mathbb{R}^{E\times Z}$, and $M = T_2(O_{B1})$, where E and Z are the width and height of the matrix m^j, respectively, and M is the set of matrix spectrograms. In particular, performing discrete statistics (taking maximum, minimum, median, standard deviation, etc.) on various PCG-transformed features for traditional classifiers is often necessary. This operation is defined as $T_3(\cdot)$ and we have $N = T_3(V + M)$. This implies discrete statistics for each slice's various vector and matrix representations, where N is the set of multifeature.
5) The characteristic expression of each slice is classified by a classifier $F(\cdot)$, for example, $c^j = F(v^j)$, $C = F(V)$; $c^j = F(m^j)$, $C = F(M)$; and $C = F(N)$, where $C = \{c^1, \dots, c^j, \dots, c^J\}$. This implies that for each slice, it is classified into a certain category. Finally, a voting mechanism is used to diagnose the positive abnormality of the whole heart sound (based on the number of normal versus abnormal slices in x^i).

B. Implementation Details

All experiments of deep learning algorithms in this article are implemented on NVIDIA Geforce RTX 3090 Ti GPUs. We use an adaptive estimation (Adam) optimizer and weight parameter optimization to train the network model in our experiments exploring the performance of different feature extraction approaches. Cross-entropy loss is used as the loss function, the learning rate is set to 0.0001, and the learning rate decay factor is set to 0.1, corresponding to decay at epochs 50 and 70. Finally, we use a tenfold cross-validation method to evaluate the performance of all approaches.

C. Heart Sound Detection Under Different Slice Lengths and Overlap Ratios

Before we discuss which features of the heart sound to use, we need to slice the heart sound to make its length computable while preserving more temporal features. Among the segmentation approaches are exact segmentation, commonly used in traditional classification models, and equal-length segmentation, employed due to the depth of the model. Since exact segmentation approaches for heart sounds are often discussed as a separate topic [32], we delve into the impact of slice length and overlap rate on heart sound detection within the context of equal-length segmentation in this section. Due to the detrimental effect of ill-conceived denoising on the classification algorithm's performance [22], in this article, we employ a sixth-order 20–950-Hz Butterworth bandpass filter [59] tailored to the heart sound frequency range. This filter uniformly preserves excess heart sound information while effectively filtering noise [53]. We choose to use VGG16 [123], Resnet18 [124], and Rsnet [125] as DNNs.

1) Slice Length: In our summary of classification approaches, the methods based on deep models [12], [36], [37], [38], [46], [61], [71], [72] no longer requires accurately segmented heart sounds. These works suggest that as long as the input sound slices contain one or more periods in length. To explore this idea, we use three datasets with heart sounds segmented into lengths of 0.25, 0.5, 0.75, 1, 1.25, and 1.5 s, without overlap, to train the three DNNs mentioned above. Since each heart sound period averages 0.8 s, we can explore the classification performance both without a complete period

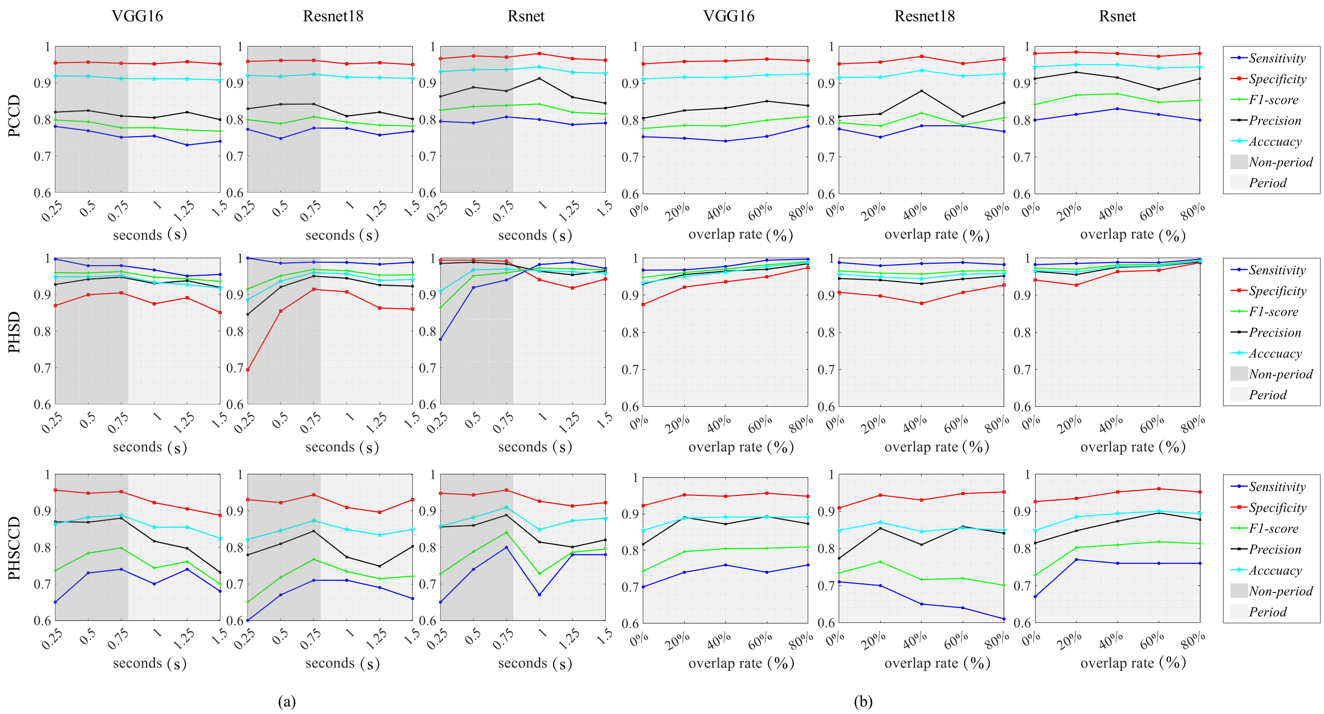

Fig. 3. Classification results under equal-length slices with different lengths and overlap rates. (a) Slice length. (b) Slice overlap rate.

and with a complete period included in the slices. As shown in Fig. 3(a), the classification metrics for different split lengths across the three datasets exhibited distinct patterns. However, the classification patterns for the three DNNs within the same dataset were essentially identical.

For the PCCD dataset, slices with incomplete periods demonstrated classification performance comparable to those with complete periods. Notably, in the VGG16 model, adding more time-domain information led to a decrease in classification performance. Therefore, we believe that DNNs can effectively utilize heart sound slices with incomplete periods.

On the PHSD dataset, the *Sensitivity* and *Specificity* of the models trained using 0.25-s slices showed a significant imbalance, with one metric being very high and the other very low. As shown in the second row of Fig. 3(a), the triangular blue line and the rectangular orange line exhibit a large discrepancy at 0.25 s. When the slice length is increased, the difference between *Sensitivity* and *Specificity* begins to decrease, resulting in a more stable model.

The experiments on the PHSCCD dataset exhibit the same pattern as those on the PHSD dataset. Specifically, the gap between *Sensitivity* and *Specificity* narrows as the slices contain more period information. However, the overall classification performance decreases when the slices consist of complete periods. Simply increasing the length of the heart sound slice does not necessarily result in a better outcome, even with minor changes. This phenomenon has been observed in other datasets as well.

These experimental results show that training the network with slices without complete periods can work as data augmentation and achieve higher classification metrics. Still, it does not always work as this. In real-world medical diagnosis, detecting abnormal heart sounds is important; meanwhile, avoiding misdiagnosing normal heart sounds is crucial, too. Misdiagnosing will lead to unnecessary further tests. Based on our experiments and experience, we recommend using DNNs with slices containing at least one complete period of the heart sound to detect abnormal heart sounds.

2) Overlap Rate: The presence of overlap in the segmentation process can effectively enhance our dataset. Overlap occurs when a sliding window is used to segment heart sounds and the window length exceeds the window shift length. This approach increases the number of training datasets by reusing information from previous slices multiple times. As the overlap rate increases, the amount of data and its utilization also increases. It is a common way to address small heart sound datasets, the scarcity of abnormal heart sounds, and the model's need for large datasets. Considering the effect of slice length on classification, the 1-s heart sound slice showed a transitional trend across the five classification metrics. This length essentially contains one complete heart sound cycle. Therefore, in this part of the study, we choose a 1-s slice length to explore further the impact of overlap rate on heart sound detection.

The experimental results are shown in Fig. 3(b). We use three models and three datasets separately. As expected, the five classification metrics show an upward trend as the overlap rate increases. The increased amount of heart sound data and the reuse of heart sound waveform features enabled the DNN to classify heart sounds more effectively, except for Resnet18, which exhibits a different trend on the PHSCCD dataset. The other models perform best at 80% overlap across all datasets, demonstrating better stability, as indicated by the smaller gap between *Sensitivity* and *Specificity*.

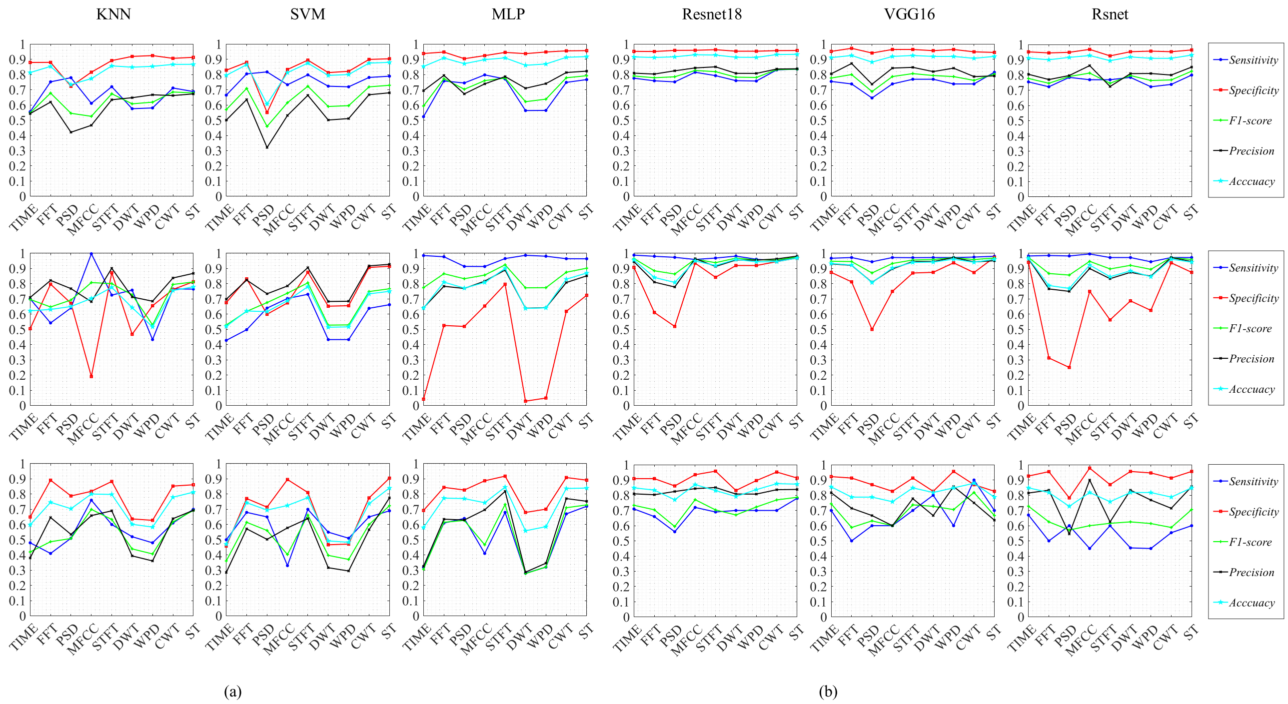

Fig. 4. Results of different feature extraction approaches on different classifiers and datasets. (a) Traditional classifier. (b) Deep learning classifier.

Similarly, different heart sound datasets have different regularities. For larger and more complex datasets (PCCD), increasing the overlap rate does not significantly enhance classification performance. The PHSD dataset contains fewer and cleaner signals than the PCCD dataset. The classification performance metrics are higher overall in the PHSD dataset, and data enhancement improves the classification results. The metrics for the same overlap rate do not differ much, and the performance is more stable. The PHSCCD dataset is less numerous and more complex than the PHSD dataset. Therefore, the overall classification result is poorer, and the performance of Resnet18 is unstable. Increasing the overlap rate in the case of equal-length segmentation is a good way to enhance the data and fully use the dataset.

Based on the analysis of this experiment results, we suggest that subsequent researchers may segment heart sound data with the same or larger overlap rate than the one presented in this article.

D. Heart Sound Detection Using Different Feature Extractions

The feature extraction approaches used in this section are all described in Section II-C. We will separately investigate the performance of features in classifying abnormal heart sounds in two kinds of classifiers. We unify the hyperparameters of the approaches to obtain them in our experiments, e.g., feature size, window function type and size, mother wavelet, decomposition level, FFT size, etc. We set the length of all slices to 1 s and the overlap rate of segmentation to 0%.

1) Under Traditional Classifiers: In this part of the experiment, we first use three traditional classifiers: KNN, SVM, and MLP. Among them, the K of KNN is set to 13. SVM employs Gaussian kernels. MLP only contains a hidden layer. Not using deep learning classifiers can eliminate the influence of the learning ability of DNNs on classification. Therefore, the traditional classifier is the best tool to verify the effect of feature extraction methods. In disease detection tasks, we tend to focus on *Sensitivity* to detect disease accurately. Still, misdiagnosis of disease can cause additional overhead and stress to healthy people, which requires more attention to *Precision*. Therefore, we combine the two and use the *F1-score* as the primary criterion for classification performance, with other metrics as auxiliary judges.

As shown in Fig. 4(a), each column represents a different classifier, and each row represents a different dataset. We can see that the best feature extraction methods are STFT, CWT, and ST. They perform highly in all datasets and classifiers, which indicates they can better preserve and highlight the distinction between normal and abnormal heart sounds. The reason is that they combine the time-domain and frequency-domain characteristics of heart sound in their features, which makes the symptoms more easily recognized by traditional classifiers. Furthermore, FFT is generally better than TIME, indicating that the pathologic information of heart sounds is more differentiated in frequency than time. However, DWT and WPD are significantly inferior to other methods. In addition, MFCC, commonly used in DNNs, performs poorly, even causing KNN to classify all the test sets in the PHSD dataset into abnormal phenomena.

2) Under Deep Learning Classifiers: In this section, we replace the three traditional classifiers with the three DNNs used in Section IV-C to compare the performance of different feature extraction approaches in DNNs.

As shown in Fig. 4(b), the performance of all feature extraction approaches is improved after using the DNN.

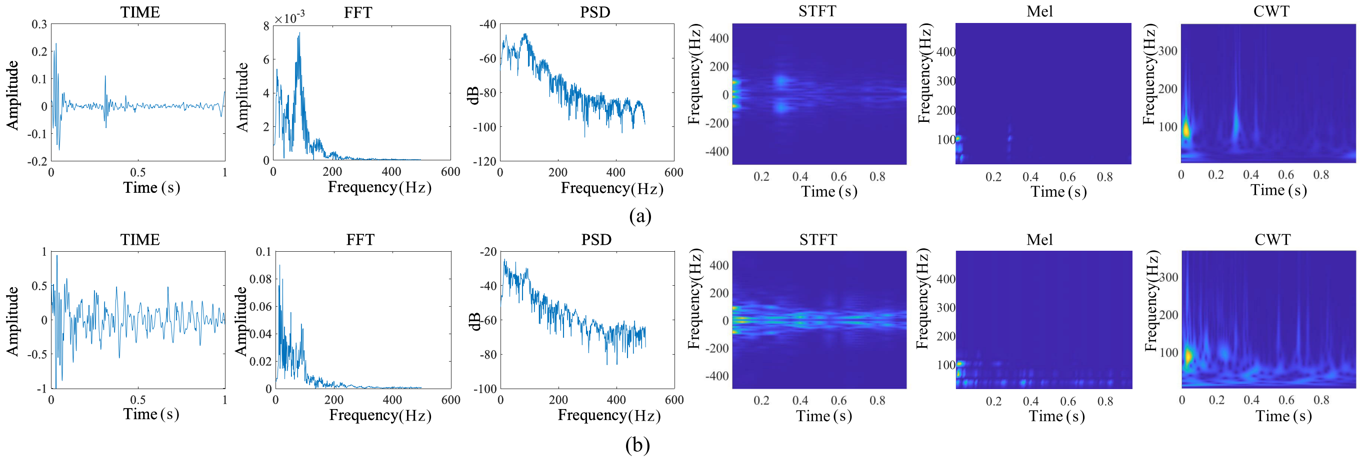

Fig. 5. Subjective effect map of the heart sound after passing the feature extraction algorithm. The first row is a segment of normal heart sound transitions and the second row is a segment of abnormal heart sound transitions. One method for each column. Please note: amplitudes are usually expressed as unitless relative values. (a) Normal. (b) Abnormal.

DNNs exhibit superior classification capabilities compared to traditional classifiers. The most obvious is the improvement of DWT and WPD feature recognition. Meanwhile, the performance of PSD in the neural network becomes poor. STFT, CWT, and ST, which perform better in traditional classifiers, still perform better in DNNs. It is worth noting that the performance of TIME and FFT feature representation methods on DNNs and traditional classifiers is the opposite. We think this is due to the convolution operation in the neural network combined with the correlation between sequential heart sounds before and after, from which the model learns features that distinguish normal and abnormal heart sounds. The time domain contains more distinguishing information about the disease symptoms.

To experience the difference between different feature extraction approaches more subjectively, we use MATLAB as a drawing tool to show the normal and abnormal heart sounds using different feature approaches. As shown in Fig. 5, we select a segment of normal and a segment of abnormal heart sounds from the dataset [7]. Normal heart sounds can be distinguished between S1 and S2 regions and have no pathological waveform interference and normal amplitude. We cannot clearly indicate the S1 and S2 regions for the abnormal heart sounds disturbed by murmurs throughout the period and exhibit abnormal amplitude and frequency in both systole and diastole. When both heart sounds are transformed with feature extraction simultaneously, the normal heart sound can have two distinct frequency components in the FFT. In contrast, the components of abnormal heart sound frequency are not different. The distinction between the two types of heart sounds is not apparent after using the PSD transform. However, three time–frequency analysis approaches (STFT, Mel, and CWT) consider both time and frequency information, which retains more information than the approaches that only consider time or frequency and facilitates the learning and classification of features by the subsequent classifier. It is noted that we do not use the spectrum of MFCC, which is a further feature extraction on the Mel spectrum, and the extracted coefficient matrix cannot visually reflect the difference between normal and abnormal heart sounds. Therefore, we just show the Mel spectrum here.

E. Reproduction Comparison

The existing heart sound detection methods use different datasets and classification performance evaluation metrics, and it is difficult to compare the performance between different approaches. We try to carry out a comparison experiment to provide a valuable analysis in this section.

To ensure a fair and comprehensive comparison, we unify the dataset and evaluation index and use the same operations for preprocessing and segmentation. During the preprocessing, we resampled the heart sounds at the sampling frequency of 2000 Hz and used the sixth-order Butterworth bandpass filter to filter. We use 1 s length and 0% overlap in the equal-length segmentation of heart sounds, and precise segmentation method [18] is used for accurate segmentation, but only on PCCD. The unification of preprocessing and segmentation also allows us to pay more attention to the feature extraction and classification stage of existing heart sound detection approaches. We use the PCCD and PHSD datasets. They satisfy the characteristics of sufficient data volume, accurate labeling, and rich heart sound information, and the exact segmentation algorithm can segment the PCCD into independent periods well, which facilitates our subsequent use of algorithms that require exact segmentation of heart sounds, especially using discrete statistical features and traditional classifier approaches. Moreover, PHSD can also be used to investigate the performance of different classification approaches on infant heart sounds.

For the compared methods, our reproduction follows: 1) for approaches with publicly available source code, we directly utilize their source code for experimentation and 2) for approaches without publicly available source code, we replicate them to the fullest extent possible based on the literature descriptions. We strive to ensure that the reproduction process remains as faithful as possible to the original approaches and maintain consistency throughout the experimental process.

TABLE IV
HEART SOUND DETECTION RESULTS BASED ON DISCRETE STATISTICAL FEATURES AND TRADITIONAL CLASSIFIERS

Methods	Seg	*Acc*	*Spe*	*Sen*	*Pre*	*F1-score*
[20]	[18]	0.7205	0.7500	**0.6897**	**0.3846**	**0.4938**
	1s	0.7256	**0.7821**	0.5489	0.3707	0.4424
	0.5s	**0.8046**	0.7500	0.5343	0.3828	0.4460
[21]	[18]	0.7950	0.8249	0.6769	0.4944	0.5714
	1s	0.8602	0.8943	0.7045	0.6388	0.6715
	0.5s	**0.8654**	**0.9064**	**0.7068**	**0.6571**	**0.6816**
[93]	[18]	**0.6739**	**0.6512**	0.7656	**0.3525**	**0.4828**
	1s	0.6708	0.6389	0.8033	0.3425	0.4803
	0.5s	0.6242	0.5545	**0.9175**	0.3255	0.4806
[94]	[18]	**0.5062**	**0.4393**	0.8103	**0.2410**	**0.3715**
	1s	0.2018	0.1190	0.9834	0.1895	0.3183
	0.5s	0.2049	0.1303	**0.9956**	0.1847	0.3188
[95]	[18]	**0.7981**	**0.8210**	**0.7077**	**0.5000**	**0.5860**
	1s	0.7357	0.7508	0.6745	0.4074	0.5086
	0.5s	0.7348	0.7781	0.5692	0.3935	0.4654
[19]	[18]	0.7640	0.7813	**0.7188**	0.4510	0.5542
	1s	0.8621	0.9066	0.7031	**0.6521**	0.6766
	0.5s	**0.8688**	**0.9179**	0.7030	0.6818	**0.6923**

We first reproduce the multifeature heart sound detection methods listed in Table I. Since this approach utilizes accurate segmentation and then extracts discrete features further from segmented heart sound slices, we use the heart sound segmentation method [18] and the PCCD dataset. Notably, the experimental results of works in Table I are missing in Table IV. The reason is that these works describe the extracted features so roughly that we cannot reproduce them exactly. We can see that Pretorius et al. [95] achieve the highest scores for *Acc*, *Pre*, and *F1-score*, while Munia et al. [94] and Homsi et al. [21] achieve the highest scores for *Sen* and *Spe*, respectively. However, the overall performance of these reproduced methods is not outstanding.

To further analyze these reproduced methods, we conduct destructive experiments. We replace the exact segmentation approaches with equal-length segmentation of 0.5 and 1 s length, which results in some statistical features (various lengths associated with heart sounds) in the constituent feature vectors being set constants. The experimental results for all methods did not uniformly decline in Table IV. Banerjee et al. [19], Potes et al. [20], and Homsi et al. [21] even outperform their previous results. However, the other methods' performance become worse. We find that these methods that do not rely on precise segmentation tend to use fewer features, with a greater emphasis on time–frequency features.

Furthermore, comparing the first row in both Fig. 4(b) and Table IV reveals that the multifeature method is less effective and robust than the direct use of DNNs. We believe these problems occur because multifeatures are obtained by further extraction and statistical computation of the features used in Section IV-D. This design operation not only relies on the experience and experimental choices of the researcher, but it is also difficult to satisfy the complexity of the heart sounds. On the other hand, traditional classifiers suffer from manual design limitations and insufficient recognition capabilities. More notably, inaccurate precise segmentation of heart sounds can also affect the subsequent classification results. Kay and Agarwal [97] demonstrate the inaccurate segmentation results of the LR-HSMM in PCCD. In summary, the multifeature heart sound detection methods require extensive extraction of various heart sound feature combinations and fine-tuning classifier parameters, often yielding unsatisfactory results. Consequently, these methods are being gradually replaced by deep learning approaches.

With the excellent performance of neural networks for pattern recognition tasks, many deep learning-based models exist for heart sound detection. Similarly, we reproduce representative deep learning-based methods for classifying heart sounds under the same dataset and conditions. The experimental results are reported in Table V. The best performance is in bold font. The second-best result is underlined. In our reproduction method, MFCC, STFT, and TIME are the characteristic expressions of heart sounds that people now pay more attention to and use. The best performance also belongs to these three feature expression methods. MFCC and STFT represent the method of combining double independent variable manual features with deep learning. TIME represents the single independent variable heart sound representation combined with deep learning, the end-to-end method. Based on the *F1-score*, MFCC performs better in PCCD, while TIME performs better in PHSD.

V. EXPERIMENTAL SUMMARY

A. Dataset

This article summarizes almost all types of heart sound detection methods according to different feature extraction approaches. We believe that it is important to have uniform datasets and evaluation indicators. According to the authors' knowledge, in addition to the datasets listed in Table III, there are also a few private datasets collected by the authors. From the experiments conducted in Figs. 3 and 4, we can conclude that experiments on datasets with different numbers and complexity will show different rules because heart sound datasets as medical datasets tend to vary from patient to patient. At the same time, various works also use different evaluation indicators. These differences make existing approaches lack reference when comparing their approaches.

B. Slice Length and Overlap Rate

Before deep learning, heart sound detection and accurate segmentation were interrelated. However, in deep learning, heart sound detection no longer requires a specially designed algorithm to segment heart sounds, only equal-length segmentation. The most important things in equal-length segmentation are segmentation length and overlap rate. The segmentation length is divided into two parts: complete period and incomplete period. To the authors' knowledge, researchers in existing heart sound detection approaches directly use different segmentation lengths containing at least one cycle and different overlap rates. In Section IV-C, we use different lengths and

TABLE V
RESULTS OF EXISTING HEART SOUND DETECTION METHODS REPRODUCED

Methods	Features	PCCD					PHSD				
		Acc	*Spe*	*Sen*	*Pre*	*F1-score*	*Acc*	*Spe*	*Sen*	*Pre*	*F1-score*
Whitaker *et al.* [103]	FFT	0.8055	0.7990	0.8125	0.6495	0.7219	0.8156	0.8045	0.8596	0.7856	0.8223
Nilanon *et al.* [61]	PSD	0.8861	0.9111	0.7889	0.6978	0.7389	0.8624	0.7492	0.9365	0.8569	0.8932
Chen *et al.* [58]	STFT	0.9070	0.9430	0.7540	0.7540	0.7540	0.8086	0.7395	0.8621	0.7219	0.7738
Karhade *et al.* [100]	FDPCT	0.8851	0.8988	0.8308	0.6750	0.7448	0.8461	0.7867	0.8726	0.8990	0.8848
Hadi *et al.* [27]	ST	0.8247	0.8284	0.8109	0.5491	0.6541	0.7355	0.6942	0.8163	0.6112	0.6866
Rubin *et al.* [70]	MFCC	0.9300	0.9550	0.8300	0.8260	0.8280	0.9230	0.8830	0.9562	0.9246	0.9415
Deng *et al.* [71]	MFCC	0.9279	0.9433	0.9125	0.9014	0.9069	0.9367	0.8857	0.9676	0.9376	0.9508
Dissanayake *et al.* [72]	MFCC	0.9254	0.8769	**0.9377**	**0.9679**	**0.9526**	0.9519	0.9735	0.9401	0.9840	0.9610
Xiao *et al.* [43] (Densenet)	TIME	0.9360	0.9570	0.8530	0.8411	0.8470	0.9624	0.9815	0.9543	0.9852	0.9695
Xiao *et al.* [43] (Cliquenet)	TIME	0.9320	0.9590	0.8300	0.8502	0.8400	0.9174	0.8925	0.9347	0.9045	0.9234
Thomae *et al.* [41]	TIME	0.8843	0.8898	0.8627	0.6761	0.7537	**0.9848**	**0.9939**	0.9709	**0.9904**	**0.9797**
Ryu *et al.* [12]	TIME	0.9068	0.9387	0.7830	0.7679	0.7747	0.9675	0.9737	0.9601	0.9487	0.9536
Tian *et al.* [46]	TIME	0.9104	0.9502	0.7543	0.7865	0.7754	0.9422	0.8814	0.9791	0.9486	0.9536
Zang *et al.* [52]	TIME	0.8996	0.9254	0.7989	0.7639	0.7701	0.8976	0.8630	0.9236	0.9218	0.9183
Fan *et al.* [53]	TIME	**0.9454**	0.9663	0.8642	0.8762	0.8644	0.9651	0.9331	0.9816	0.9697	0.9732
Maity *et al.* [66]	STFT	0.9367	0.9580	0.8145	0.8718	0.8695	0.9692	0.9466	**0.9822**	0.9704	0.9761
Ismail *et al.* [62]	STFT	0.9376	**0.9705**	0.8128	0.8783	0.8423	0.9576	0.947	0.9684	0.9661	0.9663

overlap rates to verify their effects on classification. We believe that the incomplete slices can be used in deep learning-based methods. Of course, we prefer to use full-cycle heart sound slices. Moreover, the overlap rate of segmentation can effectively enhance the data and thus improve the performance of deep learning.

C. Feature Extraction

In Section IV-D, we measure the *Accuracy*, *Sensitivity*, *Specificity*, *Precision*, and *F1-score* of SIV- and DIV-based features. Considering the high sensitivity to effectively diagnose the disease and the high precision to reduce the expense associated with misdiagnosis, our discussion is based on the *F1-score*. TIME, FFT, and PSD as SIV-based features are significant advantages in terms of computation time and memory usage. Especially, TIME without any transformation retains more information and is widely used in DNNs with superior performance; it is also an end-to-end approach because it is directly input to the model to get classification results. FFT is a good choice for using traditional classifiers. It should be realized that the SIV-based approach contains less information about the features themselves. The DIV-based features can add a dimensional perspective to represent the heart sounds. STFT, CWT, and ST outperform the other methods on the traditional classifier. For deep learning classifiers, MFCC, STFT, CWT, and ST can be learned well for classification. However, CWT and ST will take more time than MFCC and STFT.

For multifeatures, based on Section IV-E, we can see their performance often depends on the precise segmentation and traditional classifiers. The statistical features (mean, variance, maximum value, minimum value, etc.) used in multifeatures are derived from the experience of researchers and have large design limitations. Meanwhile, since the multifeature vectors are primarily short and depend on traditional classifiers for classification, the limitations of traditional classifiers are also responsible for the poor performance of multifeatures. Since there is insufficient information given in these articles on the selection of hyperparameters regarding features and classifiers, the results of our reproduced methods are not up to the ones mentioned in some of the original articles. Another reason comes from the influence of the exact segmentation algorithm. Therefore, we do not recommend trying this type of approach.

D. Reproduction Comparison

Based on our experimental results in Section IV-E, most advanced methods in recent years are based on deep learning. The best-performing methods are in two categories. One is the end-to-end approach that directly combines TIME with DNNs. Another is MFCC and STFT, which are used in the DNN. These three features are reported to perform better in Section IV-D (see Fig. 4). Therefore, discovering more suitable approaches for heart sound feature extraction and building higher performance networks are keys to future work. Moreover, since network models can feature extraction, it will be interesting to turn hand-designed features into learnable network modules and combine them with the network, making a more interpretable end-to-end approach. In addition, transferring pretrained models for image recognition to recognize heart sound spectrograms may be a good point for subsequent research.

VI. CONCLUSION

In conclusion, this article addresses which heart sound detection algorithms and features perform optimally under various conditions through a series of experiments. We also explore overlooked issues in existing research. The experiments are conducted following a standardized process encompassing preprocessing, segmentation, feature extraction, and classification. They are performed on the PCCD (large and complex), PHSD (small and high quality, focusing on children), and PHSCCD (minimal and low quality) datasets.

This study discusses the impact of segmentation length and overlap rate on subsequent classification, analyzing nine prominent heart sound features. Performance evaluations of existing heart sound detection algorithms are conducted at the same level. The experimental results indicate that during the segmentation phase, optimal performance is achieved with segment lengths that include at least one complete cycle and employ a larger overlap rate.

Regarding feature extraction, TIME and FFT features based on a single independent variable demonstrate superior performance in both deep models and conventional classifiers. STFT, CWT, and ST features based on double independent variables perform well across all classification models, while MFCC excels primarily in DNNs. The replication approach underscores that combining TIME or MFCC with DNNs currently yields the best performance. We advocate for the use of deep learning frameworks in future research endeavors.

We believe that our findings can advance heart sound detection methodologies and assist researchers in making informed decisions regarding implementation strategies.

References

[1] (2023). *World Heart Report 2023*. [Online]. Available: https://heartreport23.world-heart-federation.org/

[2] (2021). *Cardiovascular Diseases (CVDS)*. [Online]. Available: https://www.who.int/zh/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[3] Z. Jiang and S. Choi, "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 286–298, Aug. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417405002150

[4] D. S. Gerbarg, A. Taranta, M. Spagnuolo, and J. J. Hofler, "Computer analysis of phonocardiograms," *Prog. Cardiovascular Diseases*, vol. 5, no. 4, pp. 393–405, Jan. 1963. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0033062063800079

[5] 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. [Online]. Available: http://ieeexplore.ieee.org/document/8279397/

[6] J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring cardiac stress using features extracted from S1 heart sounds," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1169–1178, Apr. 2015.

[7] 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. [Online]. Available: https://iopscience.iop.org/article/10.1088/0967-3334/37/12/2181

[8] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, no. 6, p. 667, May 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/6/667

[9] 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.

[10] J. Chen, Z. Guo, X. Xu, G. Jeon, and D. Camacho, "Artificial intelligence for heart sound classification: A review," *Expert Syst.*, vol. 41, no. 4, p. 13535, Apr. 2024.

[11] T. S. Roy, J. K. Roy, N. Mandal, and S. C. Mukhopadhyay, "Recent advances in PCG signal analysis using AI: A review," *Int. J. Smart Sens. Intell. Syst.*, vol. 17, no. 1, pp. 1–37, Jan. 2024.

[12] 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.

[13] H.-L. Her and H.-W. Chiu, "Using time-frequency features to recognize abnormal heart sounds," in *Proc. Comput. Cardiology Conf. (CinC)*, Sep. 2016, pp. 1145–1147.

[14] Y. Zheng, X. Guo, and X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2710–2721, Apr. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417414006897

[15] A. A. Sepehri, A. Kocharian, A. Janani, and A. Gharehbaghi, "An intelligent phonocardiography for automated screening of pediatric heart diseases," *J. Med. Syst.*, vol. 40, no. 1, p. 16, Jan. 2016. [Online]. Available: http://link.springer.com/10.1007/s10916-015-0359-3

[16] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1746809416301616

[17] S. Yazdani, S. Schlatter, S. A. Atyabi, and J.-M. Vesin, "Identification of abnormal heart sounds," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 1157–1160.

[18] 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.

[19] R. Banerjee et al., "Time-frequency analysis of phonocardiogram for classifying heart disease," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 573–576.

[20] 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.

[21] M. N. Homsi et al., "Automatic heart sound recording classification using a nested set of ensemble algorithms," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 817–820.

[22] M. H. Asmare, F. Woldehanna, L. Janssens, and B. Vanrumste, "Can heart sound denoising be beneficial in phonocardiogram classification tasks*f*," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 354–358.

[23] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417417303305

[24] M. 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.

[25] S.-W. Deng and J.-Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Future Gener. Comput. Syst.*, vol. 60, pp. 13–21, Jul. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167739X16000121

[26] A. Bourouhou, A. Jilbab, C. Nacir, and A. Hammouch, "Detection and localization algorithm of the S1 and S2 heart sounds," in *Proc. Int. Conf. Electr. Inf. Technol. (ICEIT)*, Nov. 2017, pp. 1–4.

[27] H. M. Hadi, M. Y. Mashor, M. Z. Suboh, and M. S. Mohamed, "Classification of heart sound based on S-transform and neural network," in *Proc. 10th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, May 2010, pp. 189–192.

[28] J. P. de Vos and M. M. Blanckenberg, "Automated pediatric cardiac auscultation," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 244–252, Feb. 2007.

[29] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support vectors machine-based identification of heart valve diseases using heart sounds," *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 47–61, Jul. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260709000339

[30] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, "A robust heart sounds segmentation module based on S-transform," *Biomed. Signal Process. Control*, vol. 8, no. 3, pp. 273–281, May 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1746809412001280

[31] 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, Sep. 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S174680941100108X

[32] M. G. M. Milani, P. E. Abas, and L. C. De Silva, "A critical review of heart sound signal segmentation algorithms," *Smart Health*, vol. 24, Jun. 2022, Art. no. 100283. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352648322000186

[33] 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 Informat.*, vol. 18, no. 4, pp. 1138–1152, Jul. 2014.

[34] 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, May 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169260714000546

[35] V. Nivitha Varghees and K. I. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3861–3872, Jun. 2017.

[36] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231219304801

[37] 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.

[38] 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.

[39] G. D. Clifford et al., "Recent advances in heart sound analysis," *Physiological Meas.*, vol. 38, no. 8, pp. E10–E25, Aug. 2017. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6579/aa7ec8

[40] O. E. Badlaoui and A. Hammouch, "Discrimination between normal and heart murmurs sound, based on statistical parameters extraction and classification," in *Proc. Electric Electron., Comput. Sci., Biomed. Engineerings' Meeting (EBBT)*, Apr. 2017, pp. 1–4.

[41] 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.

[42] V. G. Sujadevi, K. P. Soman, R. Vinayakumar, and A. U. P. Sankar, "Deep models for phonocardiography (PCG) classification," in *Proc. Int. Conf. Intell. Commun. Comput. Techn. (ICCT)*, Dec. 2017, pp. 211–216.

[43] B. Xiao et al., "Follow the sound of children's heart: A deep-learning-based computer-aided pediatric CHDs diagnosis system," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1994–2004, Mar. 2020.

[44] R. Avanzato and F. Beritelli, "Heart sound multiclass analysis based on raw data and convolutional neural network," *IEEE Sensors Lett.*, vol. 4, no. 12, pp. 1–4, Dec. 2020.

[45] 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.

[46] G. Tian, C. Lian, and Z. Zeng, "Integrated Res2Net combined with seesaw loss for long-tailed PCG signal classification," in *Proc. 11th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, Dec. 2021, pp. 53–58.

[47] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169260721000146

[48] H. Li et al., "Dual-input neural network integrating feature extraction and deep learning for coronary artery disease detection using electrocardiogram and phonocardiogram," *IEEE Access*, vol. 7, pp. 146457–146469, 2019.

[49] S. Li, K. Tian, and R. Wang, "Unsupervised heart abnormality detection based on phonocardiogram analysis with beta variational auto-encoders," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 8353–8357.

[50] S. S. Hussain, M. Ashfaq, M. S. Khan, and S. Anwar, "Deep learning based phonocardiogram signals analysis for cardiovascular abnormalities detection," in *Proc. Int. Conf. Robot. Autom. Ind. (ICRAI)*, Mar. 2023, pp. 1–6.

[51] A. Patwa, M. M. U. Rahman, and T. Y. Al-Naffouri, "Heart murmur and abnormal PCG detection via wavelet scattering transform & a 1D-CNN," 2023, *arXiv:2303.11423*.

[52] J. Zang, C. Lian, B. Xu, Z. Zhang, Y. Su, and C. Xue, "AmtNet: Attentional multi-scale temporal network for phonocardiogram signal classification," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104934.

[53] J. Fan et al., "Le-LWTNet: A learnable lifting wavelet convolutional neural network for heart sound abnormality detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.

[54] R. Saraçoğlu, "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, Oct. 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0952197612001649

[55] 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. [Online]. Available: http://link.springer.com/10.1007/s10916-010-9446-7

[56] A. Yadav, M. K. Dutta, C. M. Travieso, and J. B. Alonso, "Automatic classification of normal and abnormal PCG recording heart sound recording using Fourier transform," in *Proc. IEEE Int. Work Conf. Bioinspired Intell. (IWOBI)*, Jul. 2018, pp. 1–9.

[57] L. K. P. Saputra, H. A. Nugroho, and M. Wulandari, "Feature extraction and classification of heart sound based on autoregressive power spectral density (AR-PSD)," in *Proc. 1st Int. Conf. Inf. Technol., Comput., Electr. Eng.*, Nov. 2014, pp. 139–143.

[58] Q. Chen, W. Zhang, X. Tian, X. Zhang, S. Chen, and W. Lei, "Automatic heart and lung sounds classification using convolutional neural networks," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–4.

[59] W. Zhang and J. Han, "Towards heart sound classification without segmentation using convolutional neural network," in *Proc. Comput. Cardiology (CinC)*, Sep. 2017, pp. 1–4.

[60] S. Alam, R. Banerjee, and S. Bandyopadhyay, "Murmur detection using parallel recurrent & convolutional neural networks," 2018, *arXiv:1808.04411*.

[61] T. Nilanon, J. Yao, J. Hao, S. Purushotham, and Y. Liu, "Normal / abnormal heart sound recordings classification using convolutional neural network," in *Proc. Comput. Cardiology Conf. (CinC)*, Sep. 2016, pp. 585–588.

[62] S. Ismail, B. Ismail, I. Siddiqi, and U. Akram, "PCG classification through spectrogram using transfer learning," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104075.

[63] Z. Guo, J. Chen, T. He, W. Wang, H. Abbas, and Z. Lv, "DS-CNN: Dual-stream convolutional neural networks based heart sound classification for wearable devices," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 1186–1194, Nov. 2023.

[64] M. A. A. Amal, D. Zulherman, and R. Widadi, "Klasifikasi sinyal phonocardiogram menggunakan short time Fourier transform dan convolutional neural network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 2, pp. 237–244, Apr. 2023.

[65] J. Chen et al., "A robust deep learning framework based on spectrograms for heart sound classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 21, no. 4, pp. 936–947, Aug. 2024.

[66] A. Maity, A. Pathak, and G. Saha, "Transfer learning based heart valve disease classification from phonocardiogram signal," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104805.

[67] J. J. G. Ortiz, C. P. Phoo, and J. Wiens, "Heart sound classification based on temporal alignment techniques," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 589–592.

[68] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "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. [Online]. Available: http://link.springer.com/10.1007/s10439-009-9838-3

[69] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1746809417301659

[70] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Classifying heart sound recordings using deep convolutional neural networks and Mel-frequency cepstral coefficients," in *Proc. Comput. Cardiol. Conf. (CinC)*, 2016, pp. 813–816.

[71] 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608020302306

[72] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, "A robust interpretable deep learning classifier for heart anomaly detection without segmentation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2162–2171, Jun. 2021.

[73] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals," *IEEE Access*, vol. 8, pp. 160882–160890, 2020.

[74] M. Boulares, R. Alotaibi, A. AlMansour, and A. Barnawi, "Cardiovascular disease recognition based on heartbeat segmentation and selection process," *Int. J. Environ. Res. Public Health*, vol. 18, no. 20, p. 10952, Oct. 2021. [Online]. Available: https://www.mdpi.com/1660-4601/18/20/10952

[75] J. Vepa, "Classification of heart murmurs using cepstral features and support vector machines," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 2539–2542.

[76] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection," *Comput. Biol. Med.*, vol. 100, pp. 132–143, Sep. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010482518301744

[77] A. Barnawi, M. Boulares, and R. Somai, "Simple and powerful PCG classification method based on selection and transfer learning for precision medicine application," *Bioengineering*, vol. 10, no. 3, p. 294, Feb. 2023.

[78] S. Abbas et al., "Artificial intelligence framework for heart disease classification from audio signals," *Sci. Rep.*, vol. 14, no. 1, p. 3123, Feb. 2024.

[79] 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, Dec. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417410004999

[80] H. Uğuz, "Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy," *Neural Comput. Appl.*, vol. 21, no. 7, pp. 1617–1628, Oct. 2012. [Online]. Available: http://link.springer.com/10.1007/s00521-011-0610-x

[81] S. Patidar, R. B. Pachori, and N. Garg, "Automatic diagnosis of septal defects based on tunable-Q wavelet transform of cardiac sound signals," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3315–3326, 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417414007453

[82] A. Meintjes, A. Lowe, and M. Legget, "Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 409–412. [Online]. Available: https://ieeexplore.ieee.org/document/8512284/

[83] P. Dhar, S. Dutta, and V. Mukherjee, "Cross-wavelet assisted convolution neural network (AlexNet) approach for phonocardiogram signals classification," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102142. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1746809420302895

[84] R. I. Aljohani, H. A. H. Mahmoud, A. Hafez, and M. Bayoumi, "A novel deep learning CNN for heart valve disease classification using valve sound detection," *Electron.*, vol. 12, no. 4, p. 846, 2023.

[85] G. Livanos, N. Ranganathan, and J. Jiang, "Heart sound analysis using the S transform," in *Proc. Comput. Cardiol.*, Jul. 2000, pp. 587–590.

[86] A. Moukadem, A. Dieterlen, and C. Brandt, "Shannon entropy based on the S-transform spectrogram applied on the classification of heart sounds," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 704–708.

[87] 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, Oct. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S001048251300173X

[88] K. Qian, Z. Ren, F. Dong, W.-H. Lai, B. W. Schuller, and Y. Yamamoto, "Deep wavelets for heart sound classification," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Taipei, Taiwan, Dec. 2019, pp. 1–2. [Online]. Available: https://ieeexplore.ieee.org/document/8986277/

[89] P. D. Mehr, N. J. Dabanloo, G. Attarodi, K. Maghooli, and N. Hemmati, "Diagnosis of aortic valve stenosis based on PCG signal using wavelet packet decomposition (WPD) and parametric models," in *Proc. Comput. Cardiology (CinC)*, Sep. 2017, pp. 1–4.

[90] S. Kang, R. Doroshow, J. McConnaughey, 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.

[91] S. Khaled, M. Fakhry, and A. S. Mubarak, "Classification of PCG signals using a nonlinear autoregressive network with exogenous inputs (NARX)," in *Proc. Int. Conf. Innov. Trends Commun. Comput. Eng. (ITCE)*, Feb. 2020, pp. 98–102.

[92] F. Chakir, A. Jilbab, C. Nacir, and A. Hammouch, "Phonocardiogram signals classification into normal heart sounds and heart murmur sounds," in *Proc. 11th Int. Conf. Intell. Systems: Theories Appl. (SITA)*, Oct. 2016, pp. 1–4.

[93] H. Tang, H. Chen, T. Li, and M. Zhong, "Classification of normal/abnormal heart sound recordings based on multi-domain features and back propagation neural network," 2018, *arXiv:1810.09253*.

[94] T. T. K. Munia et al., "Heart sound classification from wavelet decomposed signal using morphological and statistical features," in *Proc. Comput. Cardiology Conf. (CinC)*, Sep. 2016, pp. 597–600. [Online]. Available: http://www.cinc.org/archives/2016/pdf/172-318.pdf

[95] 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.

[96] Y. Zhang, S. Ayyar, L.-H. Chen, and E. J. Li, "Segmental convolutional neural networks for detection of cardiac abnormality with noisy heart sound recordings," 2016, *arXiv:1612.01943*.

[97] E. Kay and A. Agarwal, "DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds," *Physiological Meas.*, vol. 38, no. 8, pp. 1645–1657, Jul. 2017. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6579/aa6a3d

[98] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 613–616.

[99] M. B. Er, "Heart sounds classification using convolutional neural network with 1D-local binary pattern and 1D-local ternary pattern features," *Appl. Acoust.*, vol. 180, Sep. 2021, Art. no. 108152. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003682X21002462

[100] J. Karhade, S. Dash, S. K. Ghosh, D. K. Dash, and R. K. Tripathy, "Time-frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[101] M. Imani and H. Ghassemian, "Phonocardiography signal classification by applying feature space transformations," in *Proc. 9th Int. Symp. Telecommun. (IST)*, Dec. 2018, pp. 118–122.

[102] A. I. Humayun, S. Ghaffarzadegan, Md. 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.

[103] B. M. Whitaker and D. V. Anderson, "Heart sound classification via sparse coding," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 805–808.

[104] 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.

[105] F. Zannat, M. M. Khan, and S. Al Sohad, "Automated system for features extraction from PCG signal," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021, pp. 1–7.

[106] Y. Soeta and Y. Bito, "Detection of features of prosthetic cardiac valve sound by spectrogram analysis," *Appl. Acoust.*, vol. 89, pp. 28–33, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003682X14002266

[107] M. Faturrahman, I. Wasito, F. D. Ghaisani, and R. Mufidah, "A classification method using deep belief network for phonocardiogram signal classification," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2017, pp. 283–290.

[108] H. Coskun, Ö. Deperlioglu, and T. Yigit, "Classification of extrasystole heart sounds with MFCC features by using artificial neural network," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[109] K. Qian et al., "A bag of wavelet features for snore sound classification," *Ann. Biomed. Eng.*, vol. 47, no. 4, pp. 1000–1011, Apr. 2019. [Online]. Available: http://link.springer.com/10.1007/s10439-019-02217-0

[110] P. Qiao, Z. Yu, Z. Jingyi, and C. Zhuo, "A method for diagnosing heart sounds in adolescents based on wavelet analysis and random forest," in *Proc. Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, Jun. 2020, pp. 69–74.

[111] Y. Chen, B. Su, W. Zeng, C. Yuan, and B. Ji, "Abnormal heart sound detection from unsegmented phonocardiogram using deep features and shallow classifiers," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 26859–26883, Jul. 2023.

[112] S. Sun, "Segmentation-based adaptive feature extraction combined with Mahalanobis distance classification criterion for heart sound diagnostic system," *IEEE Sensors J.*, vol. 21, no. 9, pp. 11009–11022, May 2021.

[113] S. Ismail and B. Ismail, "PCG signal classification using a hybrid multi round transfer learning classifier," *Biocybern. Biomed. Eng.*, vol. 43, no. 1, pp. 313–334, Jan. 2023.

[114] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017.

[115] S. Bianco, M. Buzzelli, D. Mazzini, and R. Schettini, "Deep learning for logo recognition," *Neurocomputing*, vol. 245, pp. 23–30, Jul. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231217305660

[116] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, Nov. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231217308445

[117] S. Sun, B. Zhang, L. Xie, and Y. Zhang, "An unsupervised deep domain adaptation approach for robust speech recognition," *Neurocomputing*, vol. 257, pp. 79–87, Sep. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231217301492

[118] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231218303229

[119] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. *The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results*. Accessed: May 2023. [Online]. Available: http://www.peterjbentley.com/heartchallenge/index.html

[120] M. R. Mangrulkar and F. R. D. Judge. *The Michigan Heart Sound and Murmur Library*. Accessed: May 2023. [Online]. Available: https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library

[121] *Washington Heart Sound and Murmur Dataset*. Accessed: May 2023. [Online]. Available: https://depts.washington.edu/physdx/heart/demo.html

[122] *Heart Murmur Dataset Dataset That Requires Paid Access*. Accessed: May 2023. [Online]. Available: http://www.egeneralmedical.com/listohearmur.html

[123] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[124] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[125] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

Xiuli Bi received the B.Sc. and M.Sc. degrees from Shanxi Normal University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the University of Macau, China, in 2017.

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

Shizhan Tang received the B.S. and M.S. degrees from the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China, in 2021 and 2024, respectively.

His research interests include heart sound detection.

Bin Xiao received the B.S. and M.S. degrees in electrical engineering from Shanxi Normal University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Xidian University, Xi'an, in 2012.

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

Weisheng Li (Member, IEEE) 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 Chongqing University of Posts and Telecommunications, Chongqing, China. His research focuses on intelligent information processing and pattern recognition.

Xinbo Gao received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. From 2001 to 2020, he was with the School of Electronic Engineering, Xidian University. Since 2020, he has been the President of Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He has published five books and approximately 200 technical articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *International Journal of Computer Vision*, and *Pattern Recognition* in the above areas.

Prof. Gao is a fellow of the Institution of Engineering and Technology. He has served as the general chair/co-chair, the program committee chair/co-chair, and a PC member for approximately 30 major international conferences. He is on the editorial boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier).

Pietro Liò received the Ph.D. degree in complex systems and nonlinear dynamics from the Department of Engineering, School of Informatics, University of Florence, Florence, Italy, and the Ph.D. degree in theoretical genetics from the University of Pavia, Pavia, Italy, in 2007.

He is currently a Professor of computational biology with the AI Division, Department of Computer Science and Technology, University of Cambridge, Cambridge, U.K., and also a member of the Artificial Intelligence Group. He is also an Adjunct Professor with the University of Rome "La Sapienza," Rome, Italy. His research interests include graph representation learning, AI and medicine, and systems biology.

Dr. Liò is a member of Cambridge Centre for AI in Medicine, European Laboratory for Learning and Intelligent Systems (ELLIS), Academia Europaea, and Asia-Pacific Artificial Intelligence Association.