# ZCHSound: Open-Source ZJU Paediatric Heart Sound Database With Congenital Heart Disease

Weijie Jia, Yunyan Wang, Renwei Chen, Jingjing Ye, Die Li, Fei Yin, Jin Yu, Jiajia Chen, Qiang Shu, and Weize Xu, *Member, IEEE*

***Abstract*—Congenital heart disease (CHD) is a common birth defect in children. Intelligent auscultation algorithms have been proven to reduce the subjectivity of diagnoses and alleviate the workload of doctors. However, the development of this algorithm has been limited by the lack of reliable, standardized, and publicly available pediatric heart sound databases. Therefore, the objective of this research is to develop a large-scale, high-standard, high-quality, and accurately labeled pediatric CHD heart sound database. Method: From 2020 to 2022, we collaborated with experienced cardiac surgeons from three general children's hospitals to collect heart sound signals from 1259 participants using electronic stethoscopes. To ensure the accuracy of the labels, the labels for all data were confirmed by two cardiac experts. To establish the baseline of ZCHsound, we extracted 84 features and used machine learning models to evaluate the performance of the classification task. Results: The ZCHSound database was divided into two datasets: one is a high-quality, filtered clean heart sound dataset, and the other is a low-quality, noisy heart sound dataset. In the evaluation of the high-quality dataset, our random forest ensemble model achieved an F1 score of 90.3% in the classification task of normal and pathological heart sounds. Conclusion: This study has successfully established a large-scale, high-quality, rigorously standardized pediatric CHD sound database with precise disease diagnosis. This database not only provides important learning resources for clinical doctors in auscultation knowledge but also offers valuable data support for algorithm engineers in developing intelligent auscultation algorithms.**

***Index Terms*—Heart sound, database, phonocardiography (PCG), congenital heart disease (CHD).**

Manuscript received 4 May 2023; revised 2 November 2023; accepted 25 December 2023. Date of publication 9 January 2024; date of current version 19 July 2024. This work was supported by Key Research and Development Plan of Zhejiang Province under Grant 2022C03087, in part by the National Natural Science Foundation of China under Grant 82270309, in part by the Key Research and Development Plan of Hainan province under Grant ZDYF2022SHFZ121, and in part by the Natural Science Foundation of Zhejiang Province under Grant LQ22H260004. *(Weijie jia, Yunyan Wang, and Renwei Chen contributed equally to this work.) (Corresponding authors: Weize Xu; Qiang Shu.)*

Weijie Jia is with the Cardiac Center, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, China, and also with the School of Biological and Chemical Engineering, Zhejiang University of Science and Technology, China.

Yunyan Wang, Fei Yin, and Jiajia Chen are with the Binjiang Institute of Zhejiang University, China.

Renwei Chen is with the Hainan Women and Children's Medical Center, China.

Jingjing Ye and Jin Yu are with the Department of Ultrasound, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, China.

Die Li is with the Cardiac Center, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, China.

Qiang Shu and Weize Xu are with the Department of Cardiac Center, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou 310020, China (e-mail: shuqiang@zju.edu.cn; weizexu@zju.edu.cn).

Our data can be accessed and downloaded by the public at http://zchsound.ncrcch.org.cn/.

Digital Object Identifier 10.1109/TBME.2023.3348800

## I. Introduction

Congenital heart disease (CHD) is a congenital malformation caused by abnormal development of the heart and large blood vessels in the embryonic stage or failure to close the channels that should be closed after birth, which is the most common birth defect in children [1]. Common CHD includes atrial septal defect (ASD), ventricular septal defect (VSD), tetralogy of Fallot, and patent ductus arteriosus (PDA). Worldwide, the prevalence of CHD in live or term newborns is 0.94%, with an increasing trend [2]. Advancements in medical techniques have significantly improved the prognosis for patients with congenital heart disease. With timely treatment, 90% to 95% of these patients can not only survive into adulthood but also maintain a high quality of life [3]. However, there are still some children who are not diagnosed in time and miss the optimal surgery period, resulting in a series of irreversible complications [4].

Prenatal screening for CHD can diagnose fetal CHD early and recommend referral to a specialized medical institution, thereby improving the survival rate and prognosis of the child. However, prenatal diagnosis of CHD varies widely, with detection rates ranging from 40% to 60% in developed countries [5], [6]. In China, only 57.6% of the 130,000 new cases of CHD each year can be diagnosed in fetal life by fetal heart ultrasound [7]. In summary, nearly half of all children with CHD worldwide rely on neonatal screening. Currently, pulse oximetry (POX) screening is the most commonly used screening method for CHD, but this method can only detect CHD with symptoms of hypoxia [8] and the method is not applicable at the high altitude [9], [10]. Despite the development of advanced cardiac monitoring and ultrasound technologies, cardiac auscultation remains a common and most cost effective measure for first-line screening due to higher costs and varying medical conditions [11], [12].

However, cardiac auscultation is a skill that takes a long period of training to master and requires an experienced physician to perform [13]. Qiu et al. found that in China, well-trained doctors in secondary-care maternity hospitals still have poor clinical ability to diagnose CHD, and the auscultation skills of neonatal or pediatric doctors are lower than their peers in tertiary hospitals [14]. Moreover, the variability in the judgment of different physicians and the lack of quantitative indicators lead to a relatively low accuracy rate of this method [15].

With the development of artificial intelligence (AI) technology in recent years, the development of an objective and efficient computer-assisted heart sound analysis system has become a promising research area, and several research teams have proposed multiple analysis algorithms to identify abnormal heart sounds with a high accuracy rate [16], [17], [18]. Automated heart sound analysis algorithms can reduce the workload of physicians and improve the accuracy of screening [19]. However, the construction of a well-generalized heart sound analysis model requires a large amount of, labeled heart sound data for training and testing. Moreover, the robustness and accuracy of these models are highly dependent on the quality of the heart sound data. However, the number of open-access heart sound databases currently available is limited, and the data annotations included generally provide only normal or abnormal heart status of the participants, lacking more detailed disease annotation. And the existing publicly available databases have participants who are mainly adults of a wide age range, with a distinct lack of data on children's heart sounds. Heart sounds collected in adults cannot be directly used for training in the child heart sound analysis model because of the presence of non-pathological changes in children's heart sounds, including accelerated heart rate and physiological third heart sound (S3) [20].

To address these issues, this study presents the first open-source heart sound dataset for children with CHD, which has been constructed through systematic data collection and quality control measures to address the challenges associated with the lack of reliable and standardized heart sound databases for pediatric patients. The dataset comprises a significant number of high-standard, high-quality heart sound data with disease annotations and corresponding ultrasound results, providing an essential data source for researchers of intelligent auscultation algorithms both nationally and internationally. The primary objective of this study is to advance the development and evaluation of intelligent auscultation algorithms, thereby increasing the feasibility of large-scale clinical application of artificial intelligence technology in the context of CHD. Additionally, this study offers a comprehensive exploration of the applications of the pediatric CHD heart sound database, including the automatic diagnosis of CHD based on the heart sound dataset. The main contributions of this paper are presented as follows:

1) Open access database: Our open-access heart sound database is currently the largest available for children with congenital heart disease, comprising heart sound recordings from 1259 participants. The database includes 693 normal heart sound audio data and 566 heart sound audio data from patients with congenital heart disease, with participants ranging in age from 1 d to 14 years. The heart sounds were recorded using a uniform digital stethoscope at a sampling frequency of 8000 Hz, and saved in (.wav) format for accessibility.

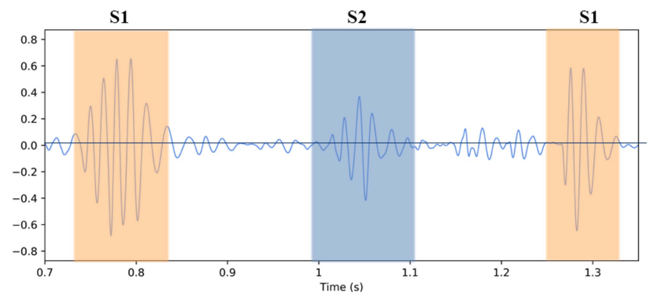



Fig. 1. Four states of the cardiac cycle from a recorded PCG(phonocardiography) signal by electronic stethoscope.

2) Label annotation method: In this study, categorical labels were assigned based on cardiac ultrasound results, which were used as the final annotation. To reduce inter-detector variability, we used a uniform acquisition device for audio recording, and uniform annotation of ultrasound results was employed to ensure the reliability of the dataset.

The rest of the paper is organized as follows. Section II describes the types of heart sounds and the current open access heart sound database. Section III provides the materials and methods of our database, including subject recruitment, participant demographics, recording tools, label annotation methods, and file naming. Section IV describes the classification effects of different classifiers on the dataset we built. Section V discusses the strengths and limitations of our database in detail and compares our database with other databases. Section VI concludes the full paper and presents our future research plans.

## II. Preliminaries

### A. Type of Heart Sound

The vibrations caused by heart activity such as contraction of the heart muscle, closure of the heart valves, and blood hitting the walls of the ventricles and aorta are transmitted through the tissues to the surface of the chest and form an acoustic signal called a heart sound [21]. These sound signals can be perceived by the human ear or recorded by electronic devices. The intensity, frequency, and correlation of heart sounds can reflect the condition of the heart valves, myocardial function, and intracardiac blood volume, and therefore the analysis of heart sounds can be used to diagnose heart disease. The heart sounds can be divided into four components according to the order in which they appear in a cardiac cycle: first heart sound (S1), second heart sound (S2), third heart sound (S3), and fourth heart sound (S4). According to the echocardiogram, it can be found that S1 occurs when the mitral and tricuspid valves go from the open state to the closed state with a duration of about 0.05 to 0.15 s and a large amplitude. S2 is mainly formed by the closure of the aortic and pulmonary valves and has a duration of 0.03–0.12 s. S3 is usually heard only in children and adolescents at a low frequency. S4 also known as the atrial sound, appears at the end of diastole and cannot be heard under normal circumstances. Fig. 1 illustrates a cardiac cycle with the locations of the fundamental heart sounds S1 and S2 marked.

### B. Existing Open-Access Heart Sound Databases

In this section, we briefly describe the relevant and publicly available heart sound datasets. We screened the datasets using the following criteria: a) data were accessible online; b) relevant to this study; and c) included information related to the recordings (including number, frequency of recording, and location of collection). Based on these criteria, three datasets were selected and we will describe them in detail. . These datasets are described in detail below.

*1) The CirCor DigiScope Dataset [22]:* This dataset contains data on heart sounds collected during the "Caravana do Coração" campaign, from 2014–2015. The database contained a total of 5282 heart sound recordings from 1568 patients, of whom 787 (50.2%) were male and 781 (49.8%) were female, with participants ranging in age from 3 days to 30 years. The heart sounds were acquired using a Littmann 3200 stethoscope from four typical auscultation points (aortic valve region, left ventricle, pulmonary valve region, and right ventricle) for 30 s at a frequency of 4000 Hz. Quality assessment of heart sound data and annotation of murmurs by two independent cardiac physiologists. The acquired audio samples were automatically segmented using three algorithms, and the segmentation results of the algorithms were examined independently by two cardiac physiologists and the mutually exclusive results were judged.

*2) Heart Sounds Shenzhen Corpus (HSS) [13]:* This dataset collected heart sound data from 170 volunteers, with participant ages ranging from 21 to 88 years old and an average age of 66 years. Heart sounds were recorded from four typical auscultation areas using an electronic stethoscope(Eko CORE, USA)equipped with Bluetooth 4.0 technology, with 30 seconds of recording for each area. The data was annotated using echocardiography, which predicted regurgitation using the area ratio of the mitral valve and tricuspid valve, classified as mild, moderate, or severe. Accordingly, the heart sound dataset has three categories: normal, mild, and moderate/severe.

*3) DigiScope2017 Dataset [23]:* The data on heart sounds were collected from 29 healthy children with participants ranging in age from 6 months to 17 years. The heart sounds were acquired at the Royal Portuguese Hospital using a Littmann 3200 stethoscope with a duration of 2–20 s. The heart sounds were acquired at 4000 Hz in the mitral position. The start and end times of S1 and S2 were manually labeled by a cardiovascular physician using software for these heart sound data.

*4) 2016 PhysioNet Challenge Dateset [24]:* This dataset is the most commonly used in recent years, with the largest amount of data and the widest range of participant ages. This data set was assembled and collated from 9 separate datasets containing a total of 2435 heart sound recordings from 1297 participants, with participants ranging in age from 18 to 86 years and acquisition durations of 8 to 312 s. Since the included datasets use different sampling devices and sampling frequencies, all audio files were re-sampled to 2000 Hz using filters and stored in. wav format. Heart sounds are collected from four different auscultation locations (aortic, pulmonary, tricuspid, and mitral), and these data include not only clean heart sounds, but also very noisy recordings.

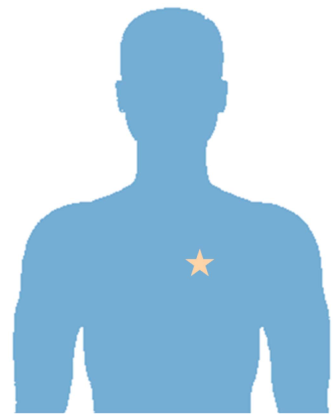


Fig. 2. Location of heart sound collection.

## III. Open-Source ZJU Paediatric Heart Sound Database With Congenital Heart Disease(ZCHSound)

### A. Subject Recruitment

The collection of heart sound data for this study was carried out at three medical institutions in China: Children's Hospital of Zhejiang University School of Medicine, Hainan Women and Children's Medical Center, and Children's Hospital of Kunming Medical University. These hospitals all have a profound pediatric professional background and rich clinical experience, providing comprehensive and meticulous health care services for children. Thus, the sample population is a good representation of the Chinese pediatric population. All participants who volunteered to participate in the study had informed consent from their parents or guardians.

### B. Sample Collection

The heart sound data used in this study were obtained using a ChildCare G-100 smart stethoscope with a sampling frequency of 8000 Hz and a quantization resolution of 16 bits. To ensure the quality and reliability of the data, the collection process was carried out by experienced physicians who were well-trained in both specimen collection and clinical measurements. To ensure consistency and validity, we collected all heart sound recordings from subjects in a supine position using a smart stethoscope. The stethoscope was placed between the second and third ribs at the left edge of the sternum for cardiac auscultation. The auscultation position is illustrated in Fig. 2. The duration of each heart sound recording ranged from 11 to 30 seconds per participant. The collected audio data were subsequently uploaded to the cloud for storage and further annotation. The specific acquisition process is illustrated in Fig. 3.

### C. Hand Corrected Signal Quality Labels

Collecting heart sounds from newborns and children poses certain difficulties, as crying, coughing, intestinal movement, and physical activity of the patients can generate additional noise. Consequently, the collected heart sound data may include a significant amount of low-quality data with noise. To facilitate

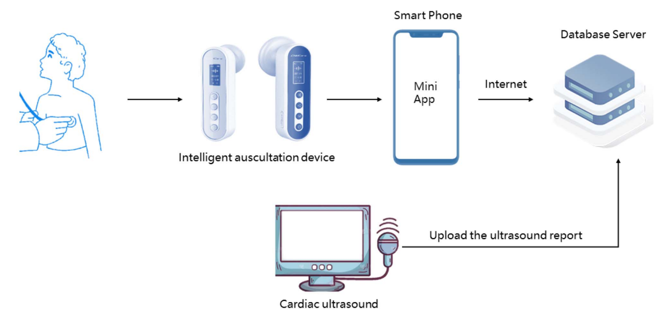



Fig. 3. Data Collection Process.

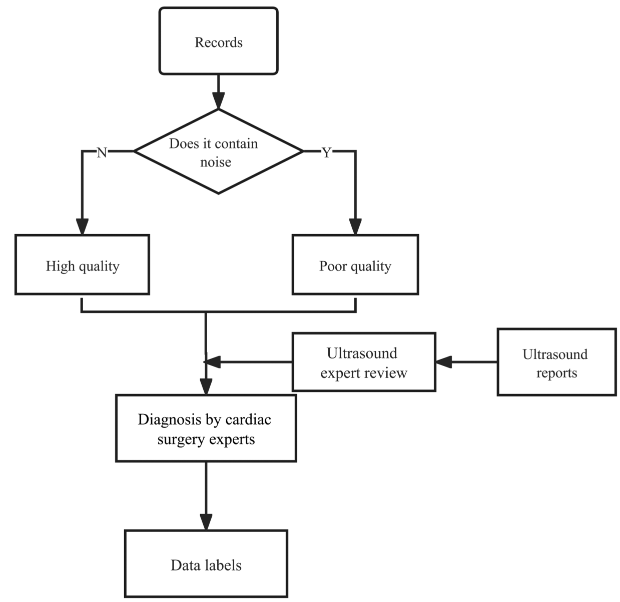



Fig. 4. Data Annotation Process.

the selection of appropriate data for analysis, we have annotated the quality of these data. Experienced physicians have screened the collected heart sound data, and we have selected noise-free data to compose a clean and high-quality heart sound dataset. The data containing noise has been used to form a noisy heart sound dataset. In clinical environments, such noise is often unavoidable and can result in errors and distortions in the heart sound signal, affecting the feature extraction and recognition of heart sound signals and reducing the accuracy of diagnosis systems. Retaining low-quality datasets can encourage researchers to identify low-quality data and perform accurate analyses. We have included heart sound data containing noise with clear noise recordings of more than 7 seconds in the low-quality heart sound dataset.

### D. Label Annotations Methodology

Accurate labels are of great importance for accurately predicting heart sound classification problems. Echocardiography can diagnose almost all heart defects [25]. In this study, all participants underwent ultrasonographic examinations conducted by an experienced sonographer, Yu. The ultrasound images and reports were subsequently stored for further analysis. These materials were meticulously reviewed by Ye, a sonographer with over two decades of experience. The final diagnoses were determined by seasoned cardiac surgeons, Shu and Xu, who based their conclusions on the cardiac recordings and ultrasound reports. These diagnoses were then utilized as data labels. The specific process is shown in Fig. 4. Fig. 5(a)–(d) show the heart sounds of a healthy child subject, a child with ventricular septal defect (VSD), a child with atrial septal defect (ASD), and a child with patent ductus arteriosus (PDA), respectively.

### E. Database Description

This study has established the first high-standard, high-quality, and disease-annotated database of pediatric CHD heart sounds, as well as the first database of newborn CHD with clinical noise and disease annotations. The high-quality heart sound dataset contains 941 participants with a total of 941 audio recordings, each approximately 20 seconds in length, resulting in a total duration of more than 5 hours. The dataset consists of 473 females (50.27%) and 468 males (49.73%), including 533 participants without heart disease serving as a control group, and the remaining patients diagnosed with various heart diseases such as atrial septal defect (ASD), patent ductus arteriosus (PDA), patent foramen ovale (PFO), ventricular septal defect (VSD), comprising 119, 32, 70, and 187 cases, respectively. The age of the participants ranged from 2 days to 14 years, with a mean age of 3 years and a median age of 8 years. We also collected 318 low-quality, noise-containing recordings of neonatal heart sounds to form a low-quality heart sound dataset, which included 318 participants, 160 of whom had no cardiac disease and served as a control group, while the remaining patients were diagnosed with cardiac diseases such as ASD, PDA, PFO, and VSD in 102, 7, 35, and 14 cases, respectively. All participants were newborns within five days of birth. According to the National Institute of Child Health and Human Development (NICHD) Pediatric Terminology classification [26], our database includes heart sound recordings from 327 neonates (26.0%), 286 infants (22.7%), 619 children (49.2%), and 27 adolescents (2.1%). Table I summarizes the detailed demographic information of our heart sound database.

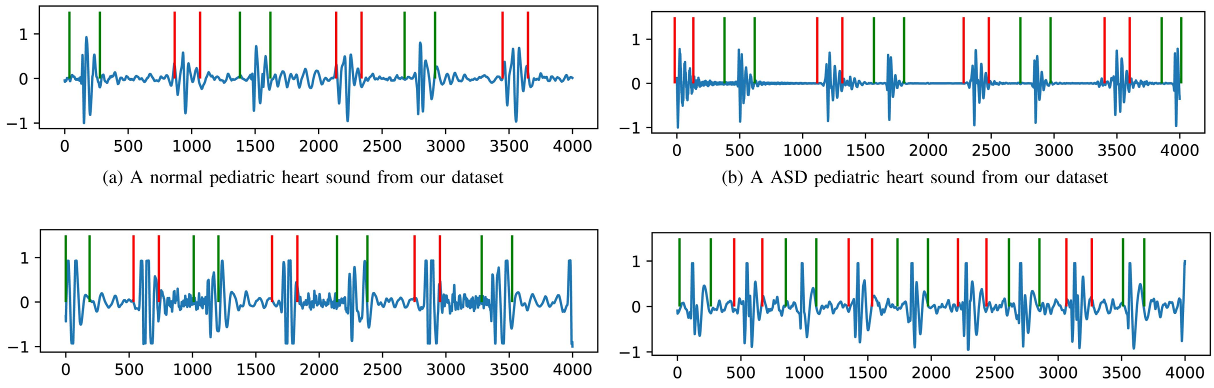
## IV. Evaluation Methods

### A. Training and Testing Sets

To assess the resilience and precision of the classification model, we employed a random hierarchical data partitioning approach in this study, dividing the dataset into a training set and a test set. Specifically, 60% of the dataset was designated as the training set, while the remaining 40% was allocated to the test set.

### B. Main Tasks

For the proposed database of children's heart sounds we propose two levels of classification tasks.




(a) A normal pediatric heart sound from our dataset

(b) A ASD pediatric heart sound from our dataset

(c) A VSD pediatric heart sound from our dataset.

(d) A PDA pediatric heart sound from our dataset.

Fig. 5. Characterization of Heart Sound Samples from Diverse Subjects: Segmenting S1 (green lines) and S2 (red lines) Boundaries with Time (Sample Points) on the x-axis and Normalized Amplitude on the y-axis. (a) A normal pediatric heart sound from our dataset. (b) A ASD pediatric heart sound from our dataset. (c) A VSD pediatric heart sound from our dataset. (d) A PDA pediatric heart sound from our dataset.

TABLE I
DEMOGRAPHIC, DIAGNOSTIC AND AGE GROUPING INFORMATION FOR PARTICIPANTS

| Variables | Total(n=1259) | High-quality heartsound database(n=941) | Low-quality heartsound database(n=318) |
|---|---|---|---|
| **Characteristics of children** | | | |
| age(years) | 2.363±2.945 | 3.160±3.015 | 0.004±0.003 |
| gender | 634(M)625(F) | 468(M)473(F) | 166(M)152(F) |
| race | Asian | Asian | Asian |
| **Primary Diagnoses** | | | |
| Normal(NOR) | 693 | 533 | 160 |
| Atrial septal defect(ASD) | 221 | 119 | 102 |
| Patent ductus arteriosus(PDA) | 39 | 32 | 7 |
| Patent foramen ovale(PFO) | 105 | 70 | 35 |
| Ventricular septal defect (VSD) | 201 | 187 | 14 |
| **Age Stages** | | | |
| neonates | 327 | 9 | 318 |
| infants | 287 | 287 | 0 |
| Toddler | 147 | 147 | 0 |
| Early childhood | 284 | 284 | 0 |
| Middle childhood | 187 | 187 | 0 |
| Early adolescence | 27 | 27 | 0 |

**Task 1 (heart sound classification, based on a clean, high-quality heart sound dataset)**

Task 1-1 involves binary classification, specifically designed to differentiate between normal heart sounds and heart sounds associated with CHD.

Tasks 1-2 is designed as multiclass classifications that aim to classify heart sounds into five categories, including normal, ASD, VSD, PDA, and PFO.

**Task 2 (heart sound classification, based on noisy low-quality heart sound dataset)**

Task 2-1 is a binary class classification designed to divide heart sounds into normal heart sounds and CHD heart sounds.

Tasks 2-2 are designed as multiclass classifications that aim to classify heart sounds into five categories, including normal, ASD, VSD, PDA, and PFO.

## C. Evaluation Metrics

To provide a standard validation criteria, we introduced the following metrics, including Accuracy (ACC), Sensitivity (SE), Specificity (SP), and F1-score (F1). They are defined as.

$$ACC(Accuracy) = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$SE(sensitivity) = \frac{TP}{TP + FN} \tag{2}$$

$$SP(specificity) = \frac{TN}{TN + FP} \tag{3}$$

$$F1 = \frac{2 \times Precision \times SE}{Precision + SE} \tag{4}$$

## D. Classification Framework for Database Quality Evaluation

This work presents a framework for the classification of high-quality (task 1) and low-quality (task 2) heart sound data, which involves three steps: preprocessing, feature extraction, and classification. The process is shown in Fig. 6. The preprocessed heart sound data is feature-extracted to generate feature vectors, which are then input into a classifier for classification. Prior studies have shown that combining feature extraction and classifiers can be effective for classification [27], [28], and efficient feature extraction methods have also led to significant results on small datasets using machine learning models [29], [30], [31]. To avoid overfitting, we chose to use a machine learning model, and we evaluated the effects of different machine learning model training approaches based on our database.

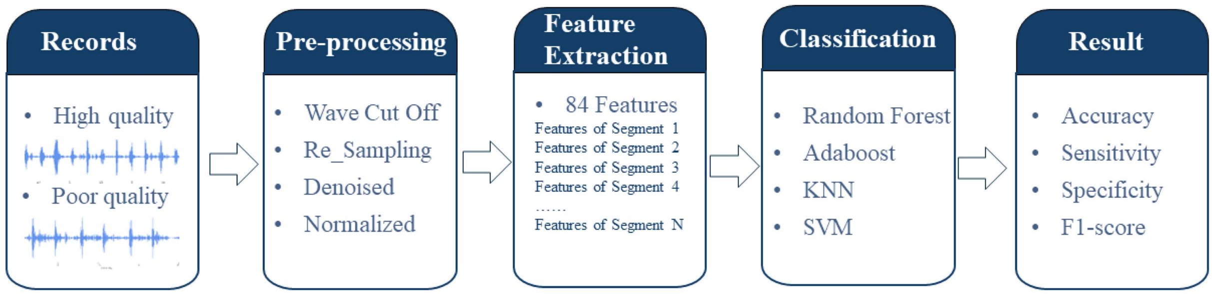



Fig. 6. Flow of sound classification framework for high-quality heart sound data and low-quality heart sound data.

TABLE II
BRIEF INTRODUCTION FOR THE 84 TIME-FREQUENCY DOMAIN FEATURES

| Feature Type | Specific Description | Feature Count |
|---|---|---|
| Duration Feature | Mean and variance of duration for each of the four cardiac phases | 8 |
| Skewness Features | Mean and variance of skewness for each of the four cardiac phases | 8 |
| Kurtosis Features | Mean and variance of kurtosis for each of the four cardiac phases | 8 |
| Short-Term Energy Features | Mean and variance of short-term energy for each of the four cardiac phases | 8 |
| Duration Ratio Features | Mean and variance of duration ratios between the S1 and systole, and the S2 and diastole | 4 |
| Mel-Frequency Cepstral Coefficients | Average of 12 MFCCs for the four cardiac phases | 48 |
| Total Features | - | 84 |

**Pre-processing Methods** The performance of classification can be improved by using signal denoising and resampling techniques [32]. During auscultation, background noise is inevitably recorded along with the heart sounds. In order to capture the complete information within heart sound signals, according to the Nyquist sampling theorem, we need a minimum sampling rate of 1000 Hz to cover the frequency range of heart sounds (20–500 Hz). However, to reduce computational resource requirements and mitigate the impact of high-frequency noise, while maintaining the integrity of the heart sound data, we have chosen to downsample the heart sound signals from 8000 Hz to 2000 Hz. This downsampling strategy, while ensuring data quality, significantly reduces data file sizes and computational resource demands. A third-order Butterworth bandpass filter with a passband of 20–650 Hz was applied to denoise the signal, retaining only the heart sounds in the frequency range of 20 Hz to 650 Hz [33]. Finally, the denoised signal was normalized to a range of $-1$ to 1.

**Feature extraction methods** In this study, we adopted the method proposed by Xu et al. [34] and combined it with relevant research to extract 84 effective temporal and spectral heart sound features based on segmented cardiac cycles for heart sound classification. Firstly, the candidates for S1 and S2 were determined by averaging normalized Shannon energy envelopes. Then, the maximum distance between adjacent candidates was used to complete the segmentation of S1 and S2 heart sound cycles. Subsequently, time and frequency features were extracted in different stages of each cardiac cycle (S1, systole, S2, and diastole).

For the time domain features, we calculated the duration, skewness, kurtosis, mean, and standard deviation of short-term energy for each stage, as well as the mean and standard deviation of the ratio of S1 to systolic pressure and S2 to diastolic pressure over the duration. These 36 features are summarized in Table II. As for spectral features, we utilized the average of 12 Mel-frequency cepstral coefficients (MFCCs) [35] for each stage, resulting in 48 spectral features of the PCG. In summary, we extracted a total of 84 features for each cardiac cycle, comprising 48 spectral features and 36 temporal features. These features can be employed to train machine learning models for heart sound classification tasks.

TABLE III
TRAINING CONFIGURATIONS

| Factors | Setting |
|---|---|
| Sample Rate | 2kHz |
| Denoising | 3-th band-pass Filter |
| Data augmentation | None |
| Feature Vectors | 84 |
| Classification Strategy | One-vs-the-rest (OvR) |

**Classification Models** Machine learning models are usually used as the main heart sound classification models with better classification results in small sample sizes [36]. Such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Adaboost we also explored the effect of the above classifiers to get the basic classification results for comparison. For a fair comparison, we use uniform parameters for model training, and the detailed configuration is shown in Table III.

## *E. Evaluation Results*

Using the scikit-learn library [37], we implemented and evaluated a classification framework for each task using machine learning models and investigated the impact of different classifiers on classification performance for a total of four models. The F1 values of each model are shown in Fig. 7.

Fig. 7 depicts the classification results of four classifiers on high-quality and low-quality heart sound data. The RF classifier yielded the highest F1 score for both Task 1-1 and Task 1-2. On the other hand, the KNN classifier showed better classification performance for Task 2-1 and Task 2-2. Based on the experimental comparison results, we used the RF and KNN models

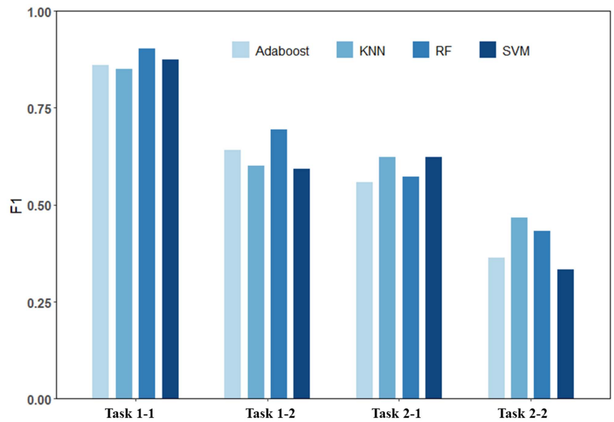



Fig. 7. Classification effects of different classifiers in task1-1, task1-2, task2-1, and task2-2.

TABLE IV
SUMMARY OF THE BEST CLASSIFICATION RESULTS FOR EACH TASK

| Model | Task | ACC | SE | SP | F1 |
|---|---|---|---|---|---|
| RF | Task 1-1 | 0.903 | 0.876 | 0.924 | 0.903 |
| RF | Task 1-2 | 0.934 | 0.555 | 1.0 | 0.695 |
| KNN | Task 2-1 | 0.625 | 0.562 | 0.687 | 0.623 |
| KNN | Task 2-2 | 0.557 | 0.25 | 0.75 | 0.466 |

for Task 1 and Task 2, respectively. As shown in Table IV, the F1 scores for Task 1-1, Task 1-2, Task 2-1, and Task 2-2 were 0.903, 0.934, 0.623, and 0.466, respectively.

Fig. 8 presents the confusion matrices for each task, where the x and y axes represent the predicted and true labels, respectively. The diagonal values represent the number of correctly identified heart sound recordings. It is observed that the best classification performance for both task 1 and task 2 is achieved for the normal class, which is the majority class, whereas the worst performance is observed for a few classes in each task.

We computed the feature importance score of all variables to identify the important features used by the model that distinguish normal and Pathological heart sounds in children. Fig. 9 provides a visual representation of the top 20 most important features employed by the model, with each feature's importance indicated on the y-axis. These variables encompass a total of 36 time-domain features and 48 frequency-domain features.

Fig. 9 shows that frequency-domain features exhibit a more significant influence on the model's output compared to time-domain features. Additionally, features associated with the systolic phase of the cardiac cycle hold a greater prominence in influencing the model's classification outcomes in comparison to diastolic phase-related features.

## V. DISCUSSIONS

In this study, we collected and annotated the ZJU Database of Heart Sounds in Children with Congenital (ZCHSound), which is currently the largest participant-based dataset of heart sounds in children with CHD, containing 1259 heart sound recordings from 1259 participants. The data was labeled based on the physician's final diagnosis. This includes four types of congenital heart diseases - ventricular septal defect, atrial septal defect, arterial ductus arteriosus, patent foramen ovale - and normal heart sounds. Compared to existing databases like CirCor DigiScope Dataset [22], Heart Sound Shenzhen Corpus [13], and PhysioNet Challenge Dateset [24], our database places a greater emphasis on classifying heart sounds in infants, children, and adolescents. The age distribution of participants in the CirCor DigiScope Dataset ranged from 0–30 years of age, the DigiScope2017 Dataset participants ranged from 0–17 years of age, the participants in the Heart Sound Shenzhen Corpus were mostly older adults with an average age of 66 years old, and the PhysioNet Challenge Dateset participants ranged from 18–86 years of age. Table V provides a detailed comparison of our database with these existing databases. PhysioNet Challenge Dataset data comes from 9 independent datasets, which means that the consistency of the data is yet to be examined; our data uses a unified acquisition device, which ensures the quality and consistency of the heart sound signals and avoids the errors generated by different devices. Furthermore, compared to other publicly available datasets, our data collection uses a higher acquisition frequency, which helps us capture the details and characteristics of heart sound signals more accurately. In addition, we retain low-quality heart sound data that contains noise so that researchers can study innovative algorithms that are more compatible with clinical use. In the data labeling process, we used a three-level labeling method. First, we categorized the data into a high-quality heart sound dataset containing clean heart sounds and a low-quality heart sound dataset containing clinical noise based on the presence of any noise. Second, all participants underwent cardiac ultrasound by an experienced sonographer. The ultrasound results and cardiac images obtained from the ultrasound were uploaded to the cloud for storage, and the ultrasound results were reexamined by an experienced sonographer. Finally, two cardiac surgeons diagnosed the patients based on the ultrasound results and the heart sound recordings, which were labeled as the disease for this heart sound recording.

Our data library offers an open-access resource for the development of heart sound classification algorithms in children with CHD. The dataset serves two primary purposes: it can be utilized to develop robust models through training on heart sound data, and it helps reduce healthcare resource wastage and disease burdens by enhancing CHD auscultation screening through the utilization of heart sound-based knowledge. Additionally, our dataset includes low-quality heart sound data, which offers a great opportunity for researchers to develop algorithms for recognizing low-quality heart sounds or reducing noise in heart sound data.

In this study, we established a comparative baseline database using machine learning models, which drew inspiration from prior research. This serves as a starting point for researchers and algorithm engineers to further develop intelligent PCG algorithms, aiding physicians in more precise diagnosis of CHD. Our experimental results indicate that the quality of PCG significantly impacts the performance of machine learning models. Therefore, in this study, we retained a dataset of low-quality heart sound data, encouraging future researchers to utilize such low-quality data to construct high-performance algorithms and enhance the robustness of the models.

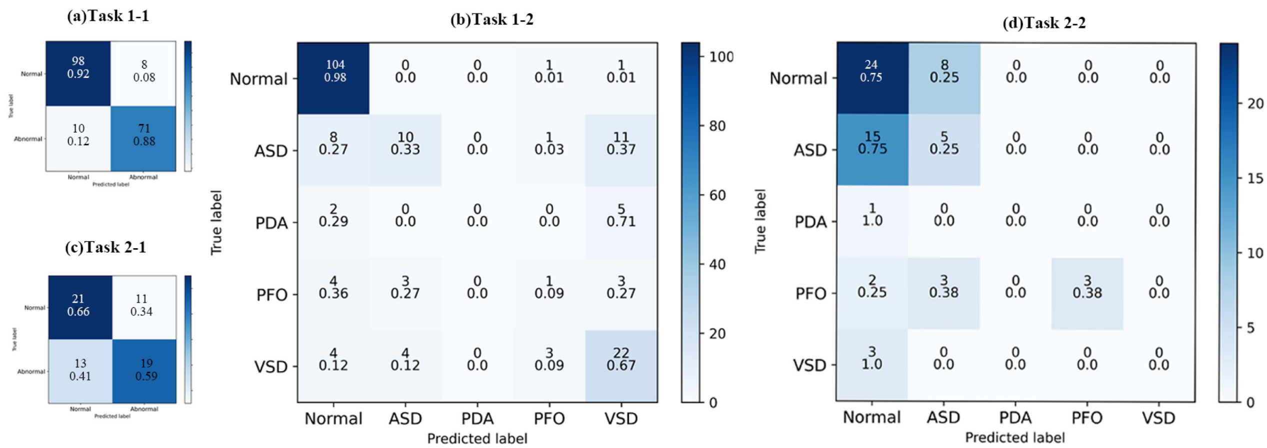



Fig. 8. Confusion matrix of Task 1-1 (a), Task 1-2 (b), Task 2-1 (c), and Task 2-2 (d).

TABLE V
SUMMARY OF EXISTING PUBLICLY AVAILABLE HEART SOUND DATASETS

| Database | Year | Sampling frequency | No. of participants | Age | Annotation categories |
|---|---|---|---|---|---|
| **Ours** | **2023** | **8000HZ** | **1259** | **0-14** | **Normal,ASD,VSD,PDA,PFO** |
| CirCor DigiScope Dataset | 2021 | 4000HZ | 1568 | 0-30 | Normal, abnormal |
| Heart Sounds Shenzhen Corpus | 2019 | 4000HZ | 170 | 21-88 | normal, mild and moderate/severe |
| DigiScope2017 Dataset | 2017 | 4000HZ | 29 | 0-17 | Normal, abnormal |
| PhysioNet Challenge Dateset | 2016 | 2000HZ | 1297 | 18-86 | Normal, abnormal |

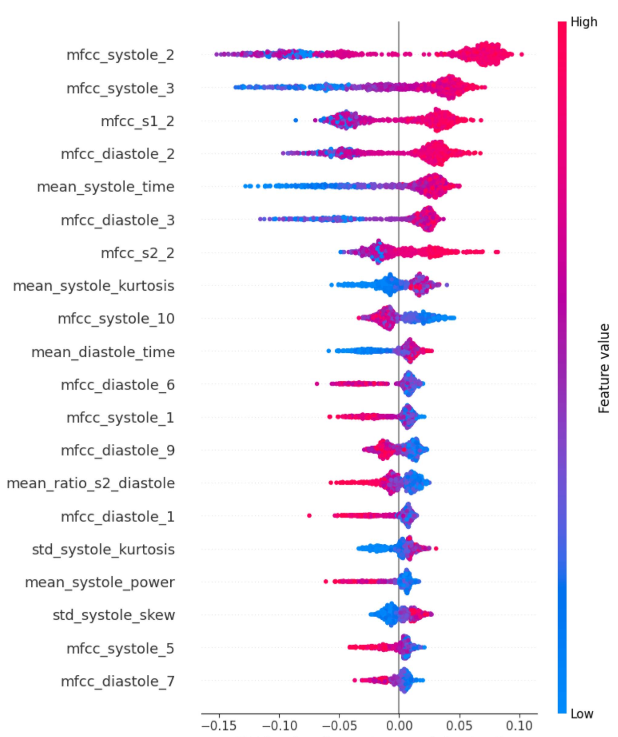



Fig. 9. Feature importance ranking based on SHapley Additive exPlanations (SHAP) values in RF model.The features have been ranked based on the cumulative sum of their SHAP values across all PCGs. SHAP values are employed to illustrate how each feature impacts the RF model outputs. In the visual representation, red signifies high feature values, while blue corresponds to low feature values. The x-axis reflects the influence of SHAP values on the model output.

Additionally, we conducted an analysis of the 84 time-frequency features mentioned in the manuscript to evaluate their significance in influencing model outputs. The results reveal that frequency-domain features hold higher importance among these features, particularly those associated with the cardiac systolic phase when compared to other phases of the cardiac cycle. This observation can be attributed to the fact that cardiac murmurs predominantly occur during the cardiac systole, thus exerting a more significant influence on the classification outcomes.

There are several limitations to our study. First, our participants are exclusively from China, which limits the generalizability of our findings. To create a more comprehensive database, we need to include participants from various regions around the world. Additionally, our heart sound data acquisition process only considers one auscultation position. While this approach improves the data's comparability and accuracy, it neglects important characteristics of other areas of the heart, leading to an incomplete evaluation of cardiac status. Lastly, there is a potential for bias in data annotation since all annotating physicians come from the same hospital with similar clinical training. To address these limitations, our next steps involve expanding the number and diversity of participants and physicians involved in our study. Furthermore, we plan to include other auscultation positions for heart sound data collection, thus enhancing the richness of our dataset.

## VI. CONCLUSION

We completed the development of an open-access pediatric CHD heart sound database (ZCHSound) containing heart sound recordings from 1259 participants. To ensure the quality of the heart sound data, we divided the dataset into two categories: clean, high-quality heart sound datasets and noisy, low-quality heart sound datasets. To ensure the reliability of data labeling, we established a more complete and comprehensive labeling process. Finally, we investigated different classifier's classification performance on these two datasets. Our data can be accessed and downloaded by the public at http://zchsound.ncrcch.org.cn/. We anticipate that our heart sound database will provide valuable support for the development of digital diagnostic tools for children with CHD and will contribute to the refinement and optimization of heart sound auscultation algorithms for this population.

Furthermore, the demand for digital-assisted diagnosis of CHD is particularly pressing in hospitals with limited healthcare resources. As such, we aim to collect and annotate heart sound screening data from multiple primary care hospitals, and subsequently make this dataset publicly available. It is our hope that the heart sound database we have created, which focuses on children with CHD, will facilitate the development of automated heart sound classification algorithms and drive the advancement of digital diagnostics for CHD. Ultimately, we believe that this work will contribute to improving the safety and well-being of children with CHD, particularly in resource-constrained settings.

## References


[1] A. C. Fahed et al., "Genetics of congenital heart disease: The glass half empty," *Circulation Res.*, vol. 112, no. 4, pp. 707–720, 2013.

[2] Y. Liu et al., "Global birth prevalence of congenital heart defects 1970–2017: Updated systematic review and meta-analysis of 260 studies," *Int. J. Epidemiol.*, vol. 48, no. 2, pp. 455–463, 2019.

[3] K. A. Holst, S. M. Said, T. J. Nelson, B. C. Cannon, and J. A. Dearani, "Current interventional and surgical management of congenital heart disease: Specific focus on valvular disease and cardiac arrhythmias," *Circ Res*, vol. 120, no. 6, pp. 1027–1044, 2017.

[4] T. van der Bom et al., "The changing epidemiology of congenital heart disease," *Nature Rev. Cardiol.*, vol. 8, no. 1, pp. 50–60, 2011.

[5] D. E. S. Jørgensen et al., "Prenatal detection of congenital heart disease in a low risk population undergoing first and second trimester screening," *Prenatal Diagnosis*, vol. 35, no. 4, pp. 325–330, 2015.

[6] C. L. van Velzen et al., "Prenatal detection of congenital heart disease–results of a national screening programme," *Int. J. Obstet. Gynaecol.*, vol. 123, no. 3, pp. 400–407, 2016.

[7] Q. M. Zhao et al., "Assessment of undiagnosed critical congenital heart disease before discharge from the maternity hospital," *Zhonghua Er Ke Za Zhi.*, vol. 55, no. 4, pp. 260–266, 2017.

[8] X.-J. Hu et al., "Pulse oximetry and auscultation for congenital heart disease detection," *Pediatrics*, vol. 140, no. 4, 2017, Art. no. e20171154.

[9] J. Wright et al., "Feasibility of critical congenital heart disease newborn screening at moderate altitude," *Pediatrics*, vol. 133, no. 3, pp. e561–569, 2014.

[10] J.-J. Li et al., "Newborn screening for congenital heart disease using echocardiography and follow-up at high altitude in China," *Int. J. Cardiol.*, vol. 274, pp. 106–112, 2019.

[11] J. M. Sztajzel et al., "Accuracy of cardiac auscultation in the era of Doppler-echocardiography: A comparison between cardiologists and internists," *Int. J. Cardiol.*, vol. 138, no. 3, pp. 308–310, 2010.

[12] X. Wang et al., "Temporal-framing adaptive network for heart sound segmentation without prior knowledge of state duration," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 650–663, Feb. 2021.

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

[14] Y.-X. Qiu et al., "Using echocardiography in newborn screening for congenital heart disease may reduce missed diagnoses," *World J. Pediatr.*, vol. 18, no. 9, pp. 629–631, 2022.

[15] I. Griebsch et al., "Comparing the clinical and economic effects of clinical examination, pulse oximetry, and echocardiography in newborn screening for congenital heart defects: A probabilistic cost-effectiveness model and value of information analysis," *Int. J. Technol. Assess. Health Care*, vol. 23, no. 2, pp. 192–204, 2007.

[16] J. Karhade et al., "Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 2506311.

[17] S. K. Ghosh et al., "Automated heart sound activity detection from PCG signal using time–frequency-domain deep neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 4006710.

[18] J. Fan et al., "Le-LWTNet: A learnable lifting wavelet convolutional neural network for heart sound abnormality detection," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 6501314.

[19] X. Huai et al., "Heart sound recognition technology based on convolutional neural network," *Inform. Health Social Care*, vol. 46, no. 3, pp. 320–332, 2021.

[20] C. Longhini et al., "The mechanism of the physiologic disappearance of the third heart sound with aging," *Jpn. Heart J.*, vol. 37, no. 2, pp. 215–226, 1996.

[21] T.-E. Chen et al., "S1 and s2 heart sound recognition using deep neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, Feb. 2017.

[22] J. Oliveira et al., "The CirCor DigiScope dataset: From murmur detection to murmur classification," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2524–2535, Jun. 2022.

[23] J. Oliveira et al., "Adaptive sojourn time HSMM for heart sound segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 642–649, Mar. 2019.

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

[25] G. Mcleod et al., "Echocardiography in congenital heart disease," *Prog. Cardiovasc Dis.*, vol. 61, no. 5, pp. 468–475, 2018.

[26] K. Williams et al., "Standard 6: Age groups for pediatric trials," *Pediatrics*, vol. 129, no. Suppl 3, pp. S153–S160, 2012.

[27] Y. N. Fuadah, M. A. Pramudito, and K. M. Lim, "An optimal approach for heart sound classification using grid search in hyperparameter optimization of machine learning," *Bioengineering*, vol. 10, no. 1, 2023, Art. no. 45.

[28] K. Iqtidar et al., "Phonocardiogram signal analysis for classification of coronary artery diseases using MFCC and 1D adaptive local ternary patterns," *Comput. Biol. Med.*, vol. 138, 2021, Art. no. 104926.

[29] A. D. Levin et al., "Extraction and assessment of diagnosis-relevant features for heart murmur classification," *Methods*, vol. 202, pp. 110–116, 2022.

[30] Y. Zheng et al.,, "A multi-scale and multi-domain heart sound feature-based machine learning model for ACC/AHA heart failure stage classification," *Physiol. Meas.*, vol. 43, no. 6, 2022, Art. no. 065002.

[31] H. Tang et al., "Automated signal quality assessment for heart sound signal by novel features and evaluation in open public datasets," *BioMed Res. Int.*, vol. 2021, 2021, Art. no. 7565398.

[32] M. Rouhani and R. Abdoli, "A comparison of different feature extraction methods for diagnosis of valvular heart diseases using PCG signals," *J. Med. Eng. Technol.*, vol. 36, no. 1, pp. 42–49, 2012.

[33] A. A. Alanazi et al., "Frequency responses of conventional and amplified stethoscopes for measuring heart sounds," *Saudi J. Med. Med. Sci.*, vol. 8, no. 2, pp. 112–117, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7305673/

[34] W. Xu et al., "Automatic pediatric congenital heart disease classification based on heart sound signal," *Art. Intell. Med.*, vol. 126, 2022, Art. no. 102257.

[35] D. M. Nogueira et al.,, "Classifying heart sounds using images of motifs, MFCC and temporal features," *J. Med. Syst.*, vol. 43, no. 6, 2019, Art. no. 168.

[36] E. B. Evangelista, F. Guajardo, and T. Ning, "Classification of abnormal heart sounds with machine learning," in *Proc. IEEE 15th Int. Conf. Signal Process.*, 2020, pp. 285–288.

[37] A. G. F. Pedregosa and G. Varoquaux, "Scikit-learn: Machine learning in Python," *Mach. Learn. Python*, vol. 12, pp. 2825–2830, 2011.