THE ANATOLIAN JOURNAL OF CARDIOLOGY

Official journal of the TURKISH SOCIETY OF CARDIOLOGY

Application of Artificial Intelligence-Based Auxiliary Diagnosis in Congenital Heart Disease Screening

ABSTRACT

Background: To evaluate the application value of artificial intelligence-based auxiliary diagnosis for congenital heart disease.

Methods: From May 2017 to December 2019, 1892 cases of congenital heart disease heart sounds were collected for learning- and memory-assisted diagnosis. The diagnosis rate and classification recognition were verified in 326 congenital heart disease cases. Auscultation and artificial intelligence-assisted diagnosis were used in 518 258 congenital heart disease screenings, and the detection accuracies of congenital heart disease and pulmonary hypertension were compared.

Results: Female sex and age > 14 years were predominant in atrial septal defect ($P < .001$) compared with ventricular septal defect/patent ductus arteriosus cases. Family history was more prominent in patent ductus arteriosus patients ($P < .001$). Compared with no pulmonary arterial hypertension, a male predominance was seen in cases of congenital heart disease–pulmonary arterial hypertension ($P < .001$), and age was significantly associated with pulmonary arterial hypertension ($P = .008$). A high prevalence of extracardiac anomalies was found in the pulmonary arterial hypertension group. A total of 326 patients were examined by artificial intelligence. The detection rate of atrial septal defect was 73.8%, which was different from that of auscultation ($P = .008$). The detection rate of ventricular septal defect was 78.8, and the detection rate of patent ductus arteriosus was 88.9%. A total of 518 258 people from 82 towns and 1220 schools were screened including 15 453 suspected and 3930 (7.58%) confirmed cases. The detection accuracy of artificial intelligence in ventricular septal defect ($P = .007$) and patent ductus arteriosus ($P = .021$) classification was higher than that of auscultation. For normal cases, the recurrent neural network had a high accuracy of 97.77% in congenital heart disease–pulmonary arterial hypertension diagnosis ($P = .032$).

Conclusion: Artificial intelligence-based diagnosis is an effective assistance method for congenital heart disease screening.

Keywords: Artificial intelligence, congenital heart disease, heart sound

INTRODUCTION

Congenital heart disease (CHD) is typically defined as a structural abnormality of the heart and/or great vessels that is present at birth. Although approximately 20% of CHD incidence can be attributed to genetic syndromes, teratogen exposure, or maternal diabetes, there remains substantial uncertainty regarding the risk factors for the remaining 80% of cases.[1]

The mean prevalence of CHD globally from 1970 to 2017 was 8.224% (7.817%, 8.641%).[2]

According to birth defect statistics, the incidence rate of CHD is 4%-50%,[3] accounting for about 30% of birth defects. With increases in the birth rate and population, the prevalence of CHD may increase by 9%-10%.[4] Twenty percent of children with CHD will progress to severe pulmonary hypertension and heart failure due to a lack of timely treatment, with high rates of death and disability, resulting in serious social and economic burdens.[5]

ORIGINAL INVESTIGATION

Hongbo Yang[1,2]
Jiahua Pan[1,2]
Weilian Wang[3]
Tao Guo[1,2]
Tengyuan Ma[1,2]

[1]*Department of Cardiology, Affiliated Cardiovascular Hospital of Kunming Medical University, Kunming, China*
[2]*Fuwai Yunnan Cardiovascular Hospital, Kunming, China*
[3]*School of Information Science and Technology, Yunnan University, Kunming, China*

Corresponding author:
Tao Guo
✉ guotao20@hotmail.com

Received: August 4, 2022
Accepted: December 8, 2022
Available Online Date: February 3, 2023

Cite this article as: Yang H, Pan J, Wang W, Guo T, Ma T. Application of artificial intelligence-based auxiliary diagnosis in congenital heart disease screening. *Anatol J Cardiol.* 2023;27(4):205-216.

DOI:10.14744/AnatolJCardiol.2022.1386

The prevalence of CHD at birth around the world has not been accurately established. Notably, the incidence rate varies, with values of 8.1%, 6.4%, 6.5% 10.3%, and 10.1% for America (1998-2005), Britain (1985-2004), Europe 22, Denmark (1977-2005), and Taiwan, respectively.[6,7] China lacks registration research. The incidence of birth defects in China in 2012 was 8.51%,[8,9] which is significantly lower than the rates in Europe and America. Moreover, information about undiagnosed CHD is difficult to obtain, especially in developing countries.[10] The variations in prevalence are primarily related to the age at diagnosis, the definition of CHD, and the screening modalities used. The most commonly used screening technology is artificial ultrasound diagnosis after auscultation. However, extensive training and experience are needed for physicians that practice auscultation.[11] It was reported that on average, just 20% of relatively inexperienced medical interns can effectively use the auscultation method to measure heart conditions.[12] Therefore, the detection rate of CHD can be greatly affected by variations in deficiencies, selection criteria, diagnostic methods, and the skills of physicians at different participating hospitals. Among the studies reporting the method of diagnosis, 95.8% utilized echocardiography, and echocardiography was not available at most facilities.

Heart sounds recorded via phonocardiogram (PCG) have been widely used in medical practice for the straightforward, efficient, and cost-effective screening of a number of cardiovascular diseases (CVD). However, the limitations highlighted in the existing studies are as follows. First, most of the previous studies ignored subject independence, which might lead to overly optimistic results. Second, only electronic stethoscope is currently used for the digital acquisition of heart sounds, and no corresponding analysis or auxiliary diagnosis method is available. To overcome the aforementioned challenges, we build a standard heart sound database with consistent data collection equipment, rigid ground truth assessment, and a reasonable data partitioning principle. We call this database the CHD Heart Sounds Yunnan (CHD-HSY) corpus. Table 1 shows the known publicly accessible heart sound databases. In the past 2 decades, building intelligent machines to monitor the status of the heart via the information extracted from PCG has become increasingly popular with developments in audio signal processing and machine learning.[13] In the era of artificial intelligence (AI) and the Internet of Things, developing an intelligent machine listening-based system can be beneficial for cardiology physicians and ultimately patients suffering from CHD by improving the general understanding of the current health status of the patients. It is encouraging to see that a variety of approaches published in the literature can be used to automatically diagnose CVD via machine learning and signal processing techniques. In this article, the application prospects of AI-assisted diagnosis methods for use in screening are discussed.

HIGHLIGHTS

- The most commonly used congenital heart disease (CHD) screening technology is cardiac auscultation, but the results depend on the physicians' skills.
- After combining phonocardiogram and electrocardiogram signals and filtering basic heart sounds, this study developed an artificial intelligence (AI)-based CHD diagnosis tool via a recurrent neural network with 3 layers.
- After screening 518 258 people, the AI detection accuracy of the ventricular septal defect ($P = .007$) and patent ductus arteriosus ($P = .021$) classification was higher than that of auscultation.

METHODS

Database

This study was a cross-sectional observational assessment. Cardiac sounds were selected from the CHD-HSY database, and the data were collected between May 2017 and December 2019 from Fuwai Yunnan Cardiovascular Hospital. More than 79 hours of recordings containing 283 800 heart sounds from 1892 (female: 1035, male: 857) subjects were obtained. The ages of the participants ranged from 1 to 72 years (10.54 ± 9.95 years).

This study was approved by the Ethics Committee of Fuwai Yunnan Cardiovascular Hospital (approval number: CHSRE2021008), and written informed consent was waived for this study given its retrospective design.

Three types of CHD cases were included: (1) atrial septal defect (ASD), (2) ventricular septal defect (VSD), and (3) patent ductal arteriosus (PDA). These 3 types are the 3 most frequent types of CHD, and these mild lesions contribute to 57.9%-65.3% of CHD cases worldwide 3. The following conditions were not included in the scope of CHD: (1) PDA that closed spontaneously within 3 months; (2) ASD <5 mm in diameter; (3) simple patent foramen ovale; (4) severe cardiovascular emergencies, such as myocardial infarction, aortic dissection, acute pulmonary embolism, etc.; (5) artificial heart valves have been implanted; (6) severe arrhythmia, ventricular rate greater than 80 beats/min; (6) severe primary valve disease and left ventricular outflow tract obstruction; and (7) patients with severe pulmonary hypertension. Echocardiography was used as the diagnostic criterion for CHD. When comparing the 2 methods in CHD screening, both suspected cases of artificial auscultation and suspected cases of AI are confirmed by echocardiography. During the development of AI, the cases of deep learning and testing are all normal or abnormal cases that have been confirmed by echocardiography.

The cardiac diagnoses were arranged in accordance with the nomenclature of the International Pediatric and Congenital Cardiac Code of the Nomenclature Working Group.[13]

Methods

In this section, we first introduce the data collection methods. Then, the acoustic features and machine learning models used for the benchmark work are defined. Finally, we present the evaluation metrics adopted in this study.

Table 1. Current Sound Databases

	Year	Number of Instances	Heart Sound Categories
PASCAL	2011	823	Normal, murmurs, extra heart sounds, and artifacts
PhysioNet	2016	3240	Normal, abnormal, and uncertain
HSS	2018	845	Normal, mild, and moderate/severe
CHD-HSY	2019	1892	ASD (770), VSD (458), PDA (664), and AI-assisted diagnosis

AI, artificial intelligence; ASD, atrial septal defect; CHD-HSY, Congenital Heart Disease—Heart Sounds Yunnan; HSS, the Heart Sounds Shenzhen;PAS, PASCAL Heart Sound Challenge Datase; PDA, patent ductal arteriosus; physioNet, PhysioNet CinC Challenge Dataset; VSD, ventricular septal defect.

Heart Sound Collecting Methods

Heart sound audio recordings were collected with an electronic stethoscope (Thinklabs One, 6500 S. Quebec Street, Suite 210, 80111 Centennial, US) with Bluetooth 4.0 at a 2 kHz sampling rate. The data were acquired from 5 locations on the body (cf. Figure 1) for 30 seconds on average (ranging from 58.6 seconds to 63.2 seconds) with the body in a supine position. In total, 9460 recordings over 79 hours were obtained from the 1892 subjects. The database included 3 categories of heart sounds to be classified: ASD, VSD, and PDA.

The system uses an STM32 lower computer, controls ad7606 to collect heart sounds and electrocardiogram (ECG) signals in parallel, and transmits the collected data to the upper computer through a W5100 network chip. The upper computer with an APSoC ZedBoard platform is divided into a Programmable Logic (PL) and a Processing System (PS). The overall design architecture of the system is shown in Figure 2.

There are 2 main research directions considered in this study: segmentation and classification. The segmentation of heart sounds separates PCG signals into their fundamental components, e.g., the first (S1) and the second (S2) heart sounds (Figure 3). S1 is caused by the closure of the mitral and tricuspid valves, and S2 is caused by the closure of the aortic and pulmonary valves. S1 and S2 are normal sounds. Nevertheless, the third and fourth heart sounds, namely S3 and S4, murmurs, and ejection clicks are usually associated with a disease or anomaly.[14]

Analysis of Heart Sounds Based on Artificial Intelligence

This study combines PCG and ECG signals to analyze the PCG results, and ECG signals are used to locate the R wave and the T wave for each heartbeat (Figure 3).[15]

Heart Sound Denoising

Heart sounds obtained using diagnostic tools are usually contaminated with noise from various sources. These sounds hinder the early detection of mild heart sounds in PCGs. Therefore, filtering noise to remove such artifacts becomes essential but should be done at the cost of preserving all diagnostic information required for PCG analysis while removing all unwanted noise. A sixth-order Chebyshev low-pass filter with a cutoff frequency of 140 Hz is used for this purpose. The noises are at high frequencies, while the diagnostic information is at low frequencies. Filtering removes the high-frequency noise. The filtered heart sound is then input into the second phase of analysis, namely segmentation.

Multiscale wavelet denoising decomposes the signal into multiple frequency scales by using a wavelet transform and processes each frequency scale separately. Wavelet transform is most suitable for nonstationary signals. Therefore, for nonstationary signals such as heart sound signals, the appropriate wavelet bases should be selected for multiscale decomposition, each scale should be separately filtered, and the filtered coefficients must be reconstructed. This process can filter noise and retain basic heart sound information.[16] Common wavelet bases include the db wavelet, coif wavelet, and sym wavelet. Figure 4 shows 3 wavelet bases: db6, coif 5, and sym5.

According to the morphological characteristics of heart sounds, the db6 wavelet is selected to filter heart sounds. However, with different levels of wavelet decomposition, the frequency distribution of each scale will vary, and the number of calculations will also differ. Therefore, selecting the appropriate level of wavelet decomposition is conducive to rapid analysis and calculation. The sampling rate of the

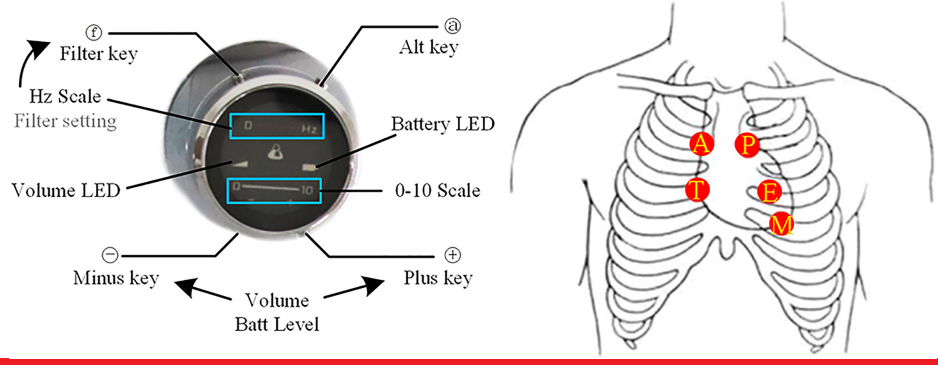

Figure 1. We collected heart sounds from 5 locations on each subject's body: auscultatory mitral (M), aortic valve auscultation (A), pulmonary valve auscultation (P), auscultatory areas of the tricuspid valve (T), and second aortic valve auscultation (E).

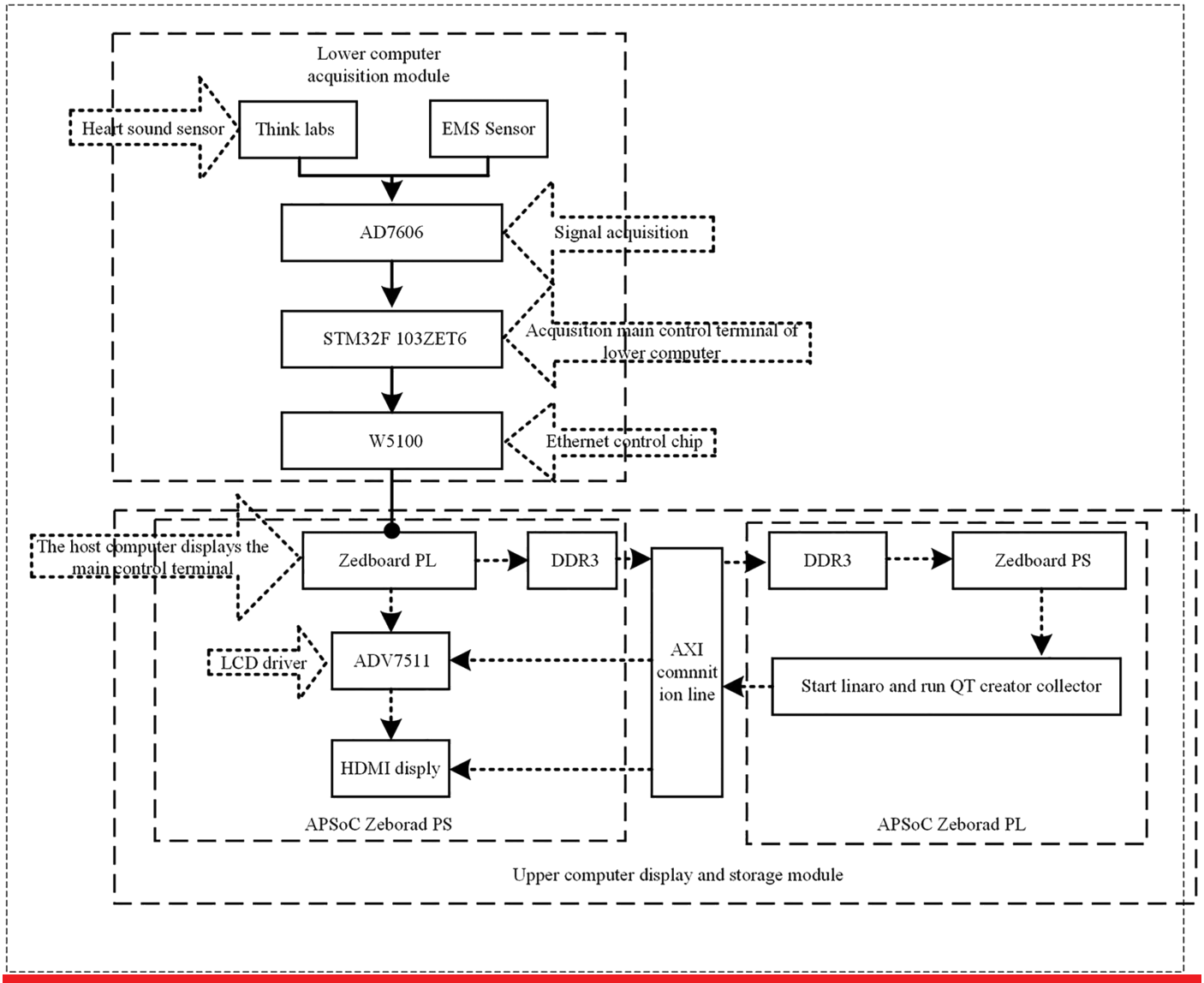

Figure 2. The overall design of the system.

heart sound acquisition box used in the subject design was 5000 Hz. According to the Nyquist sampling theorem, the maximum frequency of a heart sound signal is 2500 Hz. The heart sound signal after wavelet denoising is not involved in subsequent feature extraction and is only used in tasks involving the S1 location. Therefore, the denoising process retains the fundamental heart sounds to the greatest extent possible and filters out other sounds. Thus, a heart sound signal is decomposed into 6 layers, considering the filtering effect and calculation speed. Figure 5 shows the schematic diagram of the 6-layer decomposition.

In Figure 5, *cAi* ($i=1, 2, 3, ..., 6$) is the approximate coefficient of layer *i*, and *cDi* ($i=1, 2, 3, ..., 6$) is the detail coefficient of layer *i*. The frequency distribution of heart sounds is 5-6,00 Hz, and the main frequency components S1 and S2 are in the range of 20-1150 Hz. To effectively retain the basic heart sounds and filter other sounds, the coefficients of cA6, cD6, cD5, and cD4 are threshold filtered in multiscale filtering, and the remaining cD1-cD3 are replaced with zeroes. The threshold filtering algorithm uses soft threshold filtering. After the above wavelet coefficients are processed, they are reconstructed in turn according to the decomposition order to obtain the filtered heart sound. To compare the filtering effects of different decomposition levels on heart sounds, the db6 wavelet base is selected to compare the filtering effects of 4, 5, and 6 decomposition layers. The comparison results are shown in Figure 6.

From the experimental results, multiscale wavelet decomposition can remove the systolic and diastolic interference information, and the soft threshold filtering algorithm can retain the basic heart sounds without changing their amplitude and shape. Additionally, according to the 3 filtering effects (B), (C), and (D) in Figure 8, the deeper the number of decomposition layers, the better the filtering of interference information. Therefore, the experimental scheme of 6-layer wavelet decomposition and filtering based on the db6 wavelet are comprehensively determined through experimental analysis.

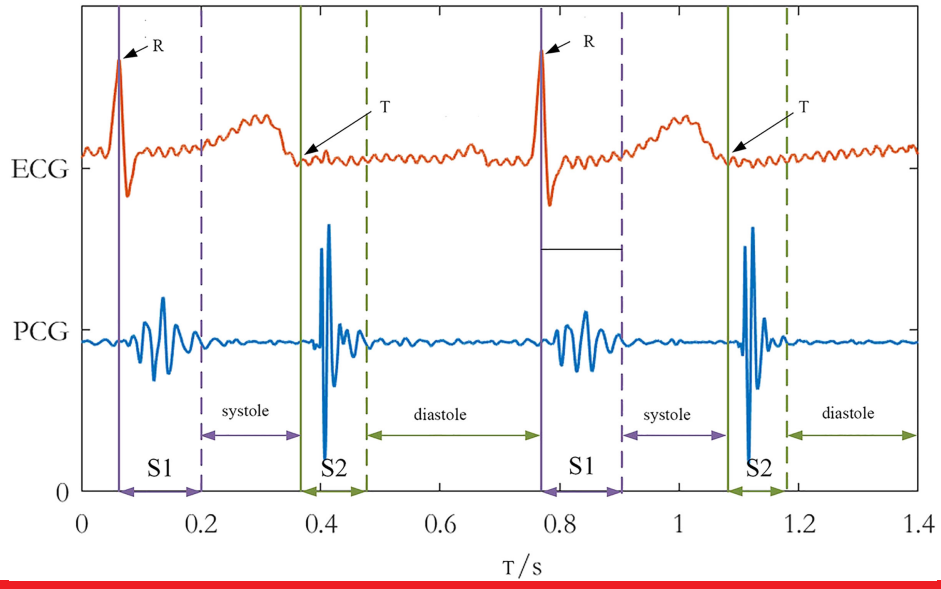

Figure 3. Heart sounds collected with the system. S1, the first heart sound; S2, the second heart sound.

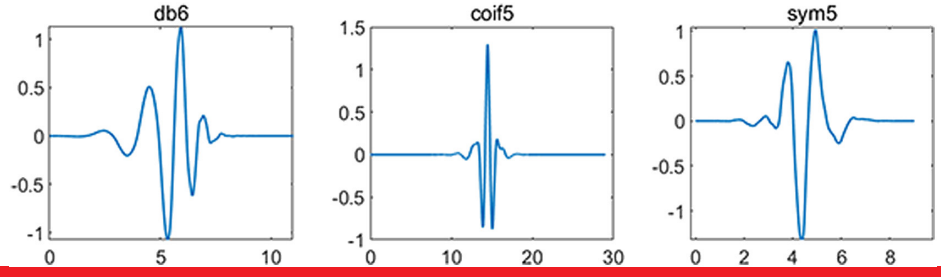

Figure 4. Comparison of 3 different wavelet bases.

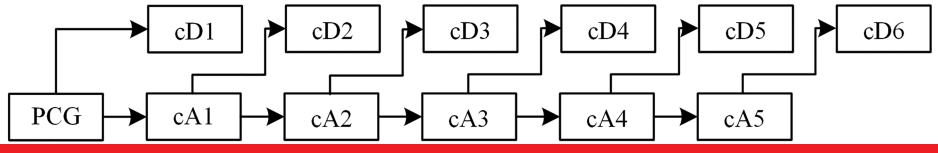

Figure 5. Schematic diagram of the 6-layer wavelet decomposition.

Segmentation of Heart Sounds

In the heart sound localization and segmentation stage, the envelope extraction of basic heart sounds is still the focus, and the integrity of basic envelope extraction directly affects the accuracy of later localization and segmentation. The extraction steps of the Shannon energy envelope with a given time scale are as follows:

1) The Shannon energy of signal X_{norm} is extracted through Eq. (1):

$$E_s = -X_{norm}{}^n log(X_{norm}{}^n) \quad (1)$$

where n is the order of Shannon energy and E_s is the Shannon energy of X_{norm}.

2) Determine the time scale L_T. The value of L_T is related to the duration of S1 and S2. According to the above calculation, the duration of S1 and S2 is 40-160 milliseconds; therefore, the value of L_T can be determined according to Eq. (2):

$$L_T = T_s \bullet F_s/2 \quad (2)$$

where T_s is 0.08 seconds or half of the longest duration of S1 and S2.

3) The mean sequence $E_s(t)$ of Shannon energy E_s is obtained according to Eq. (3):

$$E_s(t) \frac{1}{2L_T+1} \sum_{K=t-L_T}^{t+L_T} E_s(k) \quad (3)$$

where $E_s(t)$ is the mean sequence of Shannon energy E_s.

4) To eliminate the influence of signal differences, the maximum value of $E_s(t)$ is normalized and mapped to the [–1, 1] interval. Finally, the Shannon energy envelope ELs with a time scale are obtained.

Figure 7 shows the effect of extracting envelopes from the same heart sound signal by introducing the Shannon energy

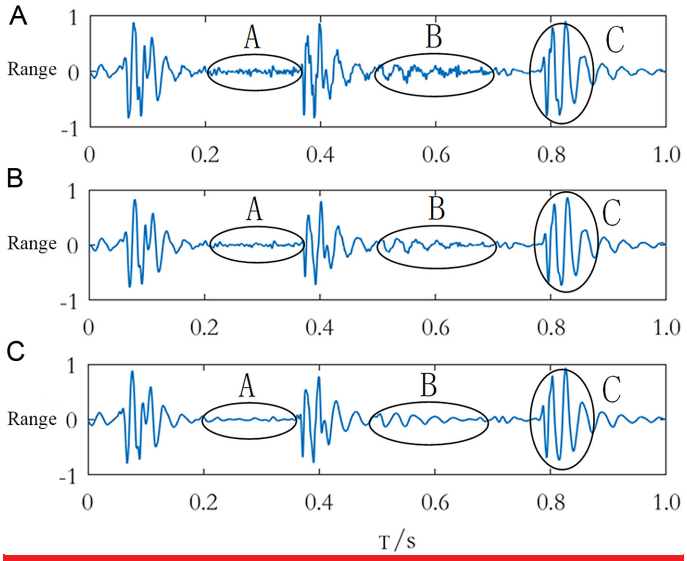

Figure 6. Comparison of multilayer wavelet denoising effects: (A) original heart sound signal, (B) 5-layer decomposition, and (C) 6-layer decomposition.

of the time scale and the normalized Shannon energy, which highlights the problem of time deviation; notably, this issue is solved by introducing the Shannon energy envelope of the time scale. Moreover, this method increases the smoothness of the envelope, and the feature retention degree of the basic heart sound envelope is basically the same as that of the normalized Shannon energy envelope.

After envelope extraction and heart sound localization, it is necessary to segment the original heart sound by using the located S1 position. However, due to the complexity of heart sounds, there is no algorithm to ensure that the success rate of heart sound localization is 100%; therefore, it is necessary to formulate different segmentation strategies for different localization situations. Additionally, to increase the temporal relevance of adjacent samples after heart sound segmentation, this article uses periodic steps or fixed steps for segmentation in time. According to the different algorithm results, 3 different segmentation strategies are formulated. Table 2 gives the corresponding segmentation methods.

A heart sound signal is a type of 1-dimensional (1D) data, and a convolutional sample (image) is an arrangement of a series of pixel values on the x-axis and y-axis. In particular, the red-green-blue values of a color image can be regarded as three 2D gray images; thus, a convolutional neural network (CNN) must organize the input samples into 2D data when performing pattern recognition. During sample training and recognition, the CNN applies a window to the input image, learns characteristic parameters in the window, and shares all connection weights. Therefore, the transformation from 1D to 2D can be realized by extracting the characteristic parameters of heart sound signals and organizing them appropriately to meet the needs of CNN recognition. In view of some new features provided by CNNs, this article uses log Mel-frequency spectral coefficients (MFSCs)[14] instead of Mel-frequency cepstral coefficients to extract heart sound signal features and organizes MFSCs into a feature map, which is

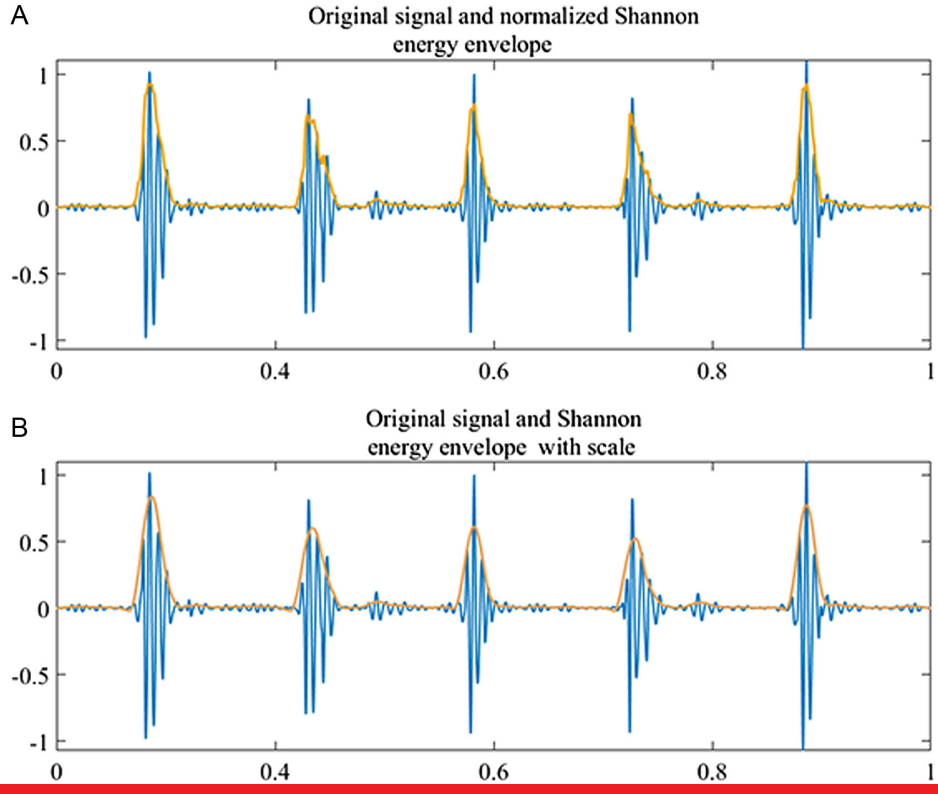

Figure 7. Comparison of two envelope extraction effects.

Table 2. Heart Sound Segmentation Strategy

Strategy No.	Judgment Mode	Division Strategy
1	S1 is greater than 1/2 of the length of the heart sound	Segmentation according to the S1 position
2	S1 is less than 1/2 of the length of the heart sound. p2p is less than 2 seconds	Starting from 0; the step value is p2p
3	S1 is less than 1/2 of the length of the heart sound. In addition, p2p is greater than 2 seconds	Split from 0; the step value is 1 s

p2p, the initial cardiac cycle length; S1, the first heart sound; S2, the second heart sound.

used as the input of the CNN in a later stage. The features extracted based on MFSCs are highly correlated with the original heart sound signal, which is smoothed in the frequency spectrum. A few MFSC features can represent most of the information in the signal.

The basic steps of MFSC processing are as follows:

1. Preprocessing. Pass the signal $x(n)$ through a high-pass filter to compensate for the high-frequency part of the signal and highlight the resonance peak at the high frequency.
2. Framing and windowing. Heart sound signals are short-term nonstationary signals that are difficult to study. Framing can make the signal approximate to a stationary signal. Windowing can prevent the signal from changing too much between 2 adjacent frames, so that there are overlapping parts between these frames, thus overcoming the phenomenon of spectrum leakage. When framing, the signal overlap is generally set to 50% to enhance the time resolution. The number of frames m in a heart sound signal after framing is calculated as shown in Eq. (4):

$$M = \frac{t \times f_s - F_l}{F_m} + 1 \tag{4}$$

In Eq. (4), t is the temporal length of the intercepted heart sound signal, F_l is the frame length, F_m is the frame shift, and f_s is the sampling frequency.

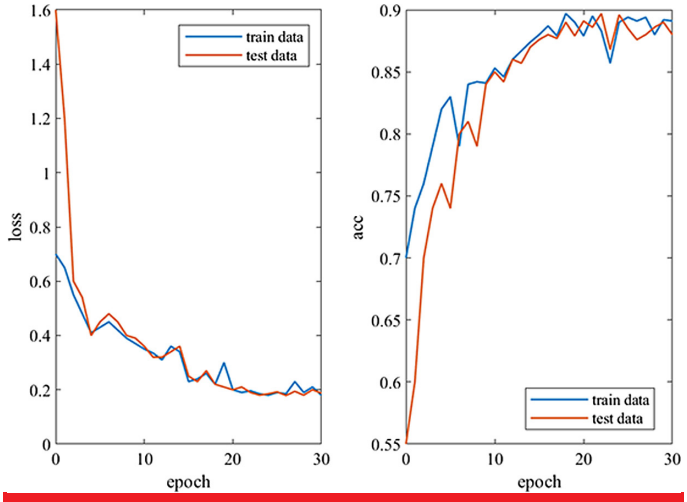

Figure 8. Adam optimizer results.

3. Frequency domain transformation. A change in the signal in the time domain is a good representation of the energy characteristics of the signal and is usually converted into an energy distribution in the frequency domain for observation. Different energy distributions represent the characteristics of different heart sound signals. In this article, Fourier transform is used, and the corresponding expression is shown in Eq. (5):

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}}, (0 \le n,k \le N-1) \tag{5}$$

where $X(k)$ represents the frequency domain signal and N is the number of points included in the frame length.

4. Mel scale conversion. The Fourier-transformed signal still needs to be converted by the Mel filter bank and Mel scale. Notably, the frequency domain signal has many redundancies. The Mel filter bank can simplify the amplitude in the frequency domain, and each frequency band is represented by a value. One filter produces a frequency band value, and the Mel scale is consistent with the auditory characteristics of human ears. Generally, the samples identified by a CNN are required to be square graphs, so the relationship between the number of filters N and the number of frames M should satisfy Eq. (6):

$$M \approx N \tag{6}$$

In Eq. (6), the value of M is generally between 24 and 80; therefore, the length of data interception can be roughly determined by Eqs (4) and (6).

5. Take the logarithm of the energy value. The human perception of sound is nonlinear, and a logarithmic nonlinear relationship can best describe the characteristics of sound perception by the human ear.
6. Determine the difference. Because a heart sound signal is continuous in the time domain, the feature information extracted from frames only reflects the characteristics of the heart sound signal in the selected frames. To make the features continuous in the time domain, the dimension information from the previous and next frames is added to the feature dimension, usually through a first-order difference (Δ) and second-order difference (Δ-Δ) scheme.

Feature Extraction and Classification of Heart Sounds

In the past few years, deep learning[17] has dominated the machine learning community, achieving high performance

in many fields, including speech recognition,[18] image recognition,[19] and object detection.[20] This approach has displayed many advantages, particularly for speech in the health domain.[21] In this study, we investigate a recurrent neural network (RNN)[22] and its performance in recognizing heart sounds. Unlike the normal feed-forward neural network, an RNN can learn contextual information from sequential inputs and consider the inherent time dependencies in heartbeat signals. In the RNN model, we use a 3-layer structure. We train the model for 1566 epochs on the training set and assess the performance of the model based on the development set at each epoch. Then, the best result is chosen and applied to the model used for the test set.

Network training refers to inputting a certain number of samples into the network and constantly optimizing the network weights through algorithm adjustment so that the output of the network is consistent with the expected values. Training a neural network mainly includes the adjustment of the network architecture, the selection of activation functions for each layer, and the selection of a model compilation optimizer. The objective is to reduce the loss value of training samples, improve the training accuracy, enhance the accuracy achieved for the verification set, and prevent overfitting. Before training, the training data are first divided, and 1000 training samples are randomly split according to a ratio of 70 : 30; that is, 700 samples are used for training, and 300 samples are used for supervised learning in each iteration. The result is verified by cross-validation. In training, the block size is 32, and the CNN model must learn 2 600 065 parameters. A logarithmic loss function is used in the supervised model, and the early stopping function provided in Keras is used to find the best model and prevent overfitting, considering the influence of different optimizer functions on the accuracy. The Adam optimizer was selected (Figure 8); the loss value was 0.25, and the accuracy rate was 0.896 after 31 rounds of model training.

In this article, a test set with 200 samples from the heart sound signal collection sample bank is used for comparative testing, and there are 100 positive samples and 100 negative samples in the test set. Classification algorithms often use sensitivity (denoted as se), specificity (denoted as sp), and accuracy (denoted as acc) as evaluation indexes. An abnormal heart sound signal that is correctly classified is called a true positive (TP), an abnormal heart sound signal that is incorrectly classified is called a false positive (FP), a correctly classified normal heart sound signal is a true negative (TN), and a normal heart sound signal that is classified as abnormal is called a false negative (FN). Specifically, se, sp, and acc are calculated as shown in Eqs (7)-(9):

$$se = \frac{TP}{TP + FN} \tag{7}$$

$$sp = \frac{TN}{TN + FP} \tag{8}$$

$$acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{9}$$

According to the prediction results of the model and the results of Eqs (7)-(9), se = 0.91, sp = 0.88, and acc = 0.895.

Comparison of Artificial Diagnosis and Artificial Intelligence Results

For 326 cases of CHD, the detection rate was classified and compared; in the screening of CHD, 518 258 children were divided into 2 groups, one group (n = 421 560) was assessed with artificial auscultation, and the other group was assessed with AI (n = 96 698) to assist with the diagnosis. Then, heart ultrasound information was used to verify and compare the accuracy of the diagnosis. To avoid missed diagnoses with AI, all children underwent artificial auscultation twice to compare the diagnostic rates of echocardiography and AI for pulmonary hypertension in patients with CHD.

Statistical Methods

For statistical description, all continuous variables were described as mean ± standard deviation, or medians and percentiles (25th percentile and 75th percentile) for non-normally distributed data. Categorical data were tested using the Pearson chi-square or Fisher's exact test. Statistical software (Stata, version 14; StataCorp, College Station, TX, USA) was used to perform all statistical analyses. A *P*-value < .05 was considered significant.

RESULTS

Data

A total of 1892 cases could be assessed with AI deep learning (54.8% males). The mean weight of individuals was 30.98 ± 15.96 kg, and the median age was 10.54 ± 9.95 years. In this study, early CHD students in primary and secondary schools were the target population for screening. Therefore, the heart sounds learned by AI mainly come from people aged less than 20 years, and the later verification population of AI is mainly hospital-confirmed patients. At the same time, in view of the fact that the incidence and treatment of CHD are mainly at low ages, the number of patients older than 20 years is relatively small (only 6.03%). In addition, in our analysis, there was no significant difference in the diagnosis rate of AI by age group. Finally, when we collect heart sounds, we keep the environment quiet, and the high sensitivity collector and AI algorithm can remove some heart sound noises, so even in the population with mixed factors, such as obese patients, we can also collect effective heart sound features for diagnosis and screening analysis.

Table 3 shows a female predominance in ASD ($P < .001$). The proportion of children (age > 14 years) was substantially higher in the ASD group than in the VSD/PDA group. Family history was appreciably predominant in the patients with PDA ($P < .001$). Compared with that for no pulmonary arterial hypertension (PAH), male predominance was observed in cases of CHD-PAH ($P < .001$), and age was significantly associated with PAH ($P = .008$). A comparatively higher prevalence of extracardiac anomalies was found in the PAH group. In total, 8 CHD-PAH patients (2.14%) had a family history. There was no significant difference between the training and test results in the general case.

Table 3. Number of Instances in Each Data Set of the CHD-HSY Corpus

CHD (n = 1892)					
	ASD (n = 770)	VSD (n = 664)	PDA (n = 458)	Train/Test (1566/326)	CHD-PAH (n = 373)
Males/females, n (%)	250 (32.49)[1,2]	307 (46.27)	221 (48.37)	678 (43.29)/146 (44.78)	217 (58.17)*
Age > 14 years, n (%)	525 (68.18)[1,2]	332 (50%)	262 (57.2)	442 (28.22)/96 (29.44)	226 (60.58)*
Extracardiac anomalies, n (%)	48 (6.23)	46 (6.92)	62 (13.53)	76 (4.98)/15 (4.6)	66 (17.69)*
Family history, n (%)	17 (2.2)[2]	28 (4.21)[2]	46 (10.04)	12 (7.66)/9 (2.76)	8 (2.14)

[1]$P < .05$ compared with VSD; [2]$P < .05$ compared with PDA; [3]$P < .05$ compared with the T test; *$P < .05$ compared with No-PAH.
ASD, atrial septal defect; CHD-HSY, Congenital Heart Disease Heart Sounds Yunnan; PAH, pulmonary arterial hypertension; PDA, patent ductal arteriosus; VSD, ventricular septal defect.

Table 4. Prevalence and Percentages of CHD Screening

CHD Subtype	Prevalence of CHD Subtype Per Thousand, n (%)	Percentage of CHD Subtype, %
Ventricular septal defect	1238 (2.39)	31.5
Atrial septal defect	678 (1.31)	17.25
Patent ductus arteriosus	388 (0.75)	9.87

CHD, congenital heart disease.

From January to December 2019, 518 258 people were screened using the study protocol; 5 prefectures (Fuyuan County, Xuanwei City, Baoshan City, Diqing state, and Wenshan state) in Yunnan and 5 provinces (Battambang, Siem Reap, Kampong Tong, Kampong Cham, and Tebenkmon) in Cambodia were included in the screening. A total of 518 258 people from 82 towns and 1220 schools were screened, and these individuals included 15 453 suspected and 3 930 (7.58%) confirmed cases. These cases included 2304 (58.62%) with a single defect. Table 4 shows that the most common subtype of CHD was VSD (31.5%), followed by ASD (17.25%), PDA (79.87%), PS (4.22%), and tetralogy of Fallot (4.7%).

There was no significant difference between the prevalences estimated in different regions and different nations (Table 4). When the analysis was restricted to ASD only, the prevalence in Yunnan was significantly higher than that in Cambodia.

Comparison of Discovery Rates

Among the 326 confirmed CHD patients, including ASD (126), VSD (137), and PDA (63) patients, the detection rate of AI diagnosis was ASD (78.6%), VSD (81.02%), and PDA (81%), as shown in Table 5. Table 6 shows that only 68 patients were correctly diagnosed with ASD diagnosis, and there was a significant difference between the 2 diagnosis types ($P = .008$). In the validation of patients with confirmed CHD, we statistically analyzed and compared the correct diagnostic rate of AI based on the age of 14 years and found that there was no significant difference in the diagnostic accuracy of ASD, VSD, and PDA (Table 7).

Accuracy Comparison

Among the 518 258 people screened, 13 126 and 3120 suspected cases were obtained in the artificial auscultation group and the AI group, respectively. Table 8 shows that the detection accuracy of AI in VSD and PDA classification was higher than that of artificial auscultation, and the difference was statistically significant. Additionally, the overall accuracy of AI screening was higher than that of auscultation screening.

Application of Artificial Intelligence-Assisted Diagnosis in Congenital Heart Disease—Pulmonary Arterial Hypertension

Congenital heart disease—pulmonary arterial hypertension heart sounds are significantly different from normal heart sounds and can be distinguished and classified by pattern recognition and AI technology. For normal cases, the RNN ($P = .032$) exhibited a high accuracy of 97.77% in CHD-PAH diagnosis (Table 9).

DISCUSSION

The system uses the APSoC-ZedBoard platform as the core and combines it with STM32 at the PL end to realize the real-time parallel acquisition of heart sounds and ECG signals. A customized embedded operating system, Linaro, is built on

Table 5. Detection Rate

	Auscultation, n (%)	AI, n (%)
ASD ($n = 126$)	99 (78.6)	93 (73.8)
VSD ($n = 137$)	111 (81.02)	108 (78.8)
PDA ($n = 63$)	51 (81)	56 (88.9)

AI, artificial intelligence; ASD, atrial septal defect; PDA, patent ductal arteriosus; VSD, ventricular septal defect.

Table 6. Number of ASD Cases Detected

Auscultation	Correct	Incorrect	*P*
AI			
Correct	68	31 (31.3)	.008
Incorrect	25	2	
All	93	33	

AI, artificial intelligence; ASD, atrial septal defect.

Table 7. Age Group Comparison

	ASD (n = 126)		VSD (n = 137)		PDA (n = 63)	
Age (years)	**AI Error**	**AI Correct**	**AI Error**	**AI Correct**	**AI Error**	**AI Correct**
≤14	20 (21.7)	63 (61.3)	20 (19.5)	72 (72.5)	4 (3.9)	31 (31.1)
>14	13 (11.3)	30 (31.7)	9 (9.5)	36 (35.5)	3 (3.1)	25 (24.9)
	$\chi2 = 0.5517$, $P = .458$		$\chi2 = 0.0548$, $P = .815$		*$P = 1.000$	

The number in the bracket is the expected value. *For expected values <5, Fisher's exact methods is used.
AI, artificial intelligence; ASD, atrial septal defect; PDA, patent ductal arteriosus; VSD, ventricular septal defect.

Table 8. Accuracy Comparison

	Auscultation Suspected Cases (n = 23 126)	AI Suspected Cases (n = 3120)	*P*
Ultrasound, n (%)	3162 (13.67)	768 (24.61)	.047
ASD, n (%)	537 (17.16)	117 (15.21)	.790
VSD	1011 (32.21)	326 (42.50)	.007
PDA	316 (10.13)	188 (24.50)	.021

AI, artificial intelligence; ASD, atrial septal defect; PDA, patent ductal arteriosus; VSD, ventricular septal defect.

Table 9. Signal Classification and Recognition Based on the RNN

Sounds	TP	TN	FP	FN	Se (%)	Sp (%)	Acc (%)
Normal	36	44	0	8	81.81	100	90
CHD-PAH	44	36	8	0	100	81.81	90
Total	80	80	8	8	90	90	90

CHD-PAH, congenital heart disease-associated pulmonary arterial hypertension; FN, false negative; FP, false positive; RNN, recurrent neural network; TN, true negative; TP, true positive.

the PS, and it runs the acquisition program compiled by Qt, creates waveforms, and stores ECG signal data in real-time. The system is portable and stable, and it provides antijamming and visualization capabilities, making it convenient for professionals to apply in the diagnosis of pathological information for patients with CHD. The system combines the acquisition part of the lower computer, the data processing part of the upper computer, and the display part and the power supply part in a polyethylene plastic box at a low cost. In this article, a heart sound signal classification algorithm based on a CNN is proposed. First, a 1D heart sound signal with considerable background noise is converted into a clean 2D sample map by using the preprocessing algorithm proposed in this article to meet the needs of the CNN classifier. These feature maps are used as CNN inputs to optimize and train the network architecture. Experiments show that the Adam optimizer provides high accuracy in the 5-layer CNN model, and the training results include a model training accuracy of 0.896 and a loss of 0.25. The algorithm yields se = 0.91, sp = 0.88, and acc = 0.895 for the test dataset. The accuracy achieved in this article is competitive with that of other algorithms. The model is optimized under the condition of large samples, and its universality and robustness are guaranteed. Compared with the methods used by traditional algorithms, the developed preprocessing method is relatively simple. The algorithm in this article can provide an excellent choice for future applications involving machine-assisted auscultation and telemedicine. In the future, we will focus on improving the accuracy and practical application of the algorithm.

Increased use of echocardiography and improved techniques are likely to increase the prevalence of ASD diagnosis, but economic conditions may be limiting. Although many people and institutions have begun to monitor birth defects, for several reasons, it is difficult to obtain convincing epidemiologic data regarding CHD prevalence using auscultation.[23] First, the screening for CHD can be greatly affected by variations in environmental noise, diagnostic methods, and the skills of physicians in different regions. Second, the narrow window (mostly within 7 days after delivery) usually makes it difficult to identify CHD with symptoms that develop after discharge.[24,25] Third, echocardiographic analysis is not available at most facilities, and this approach is not suitable for large-scale screening. Given these factors, the surprisingly low prevalence of CHD (1.7-5.2 per 1000 people) derived from surveillance system data in China is understandable.[8,26]

The accuracy of our screening method has been proven to be satisfactory. More importantly, AI-assisted diagnosis is used in screening. Extensive follow-up modalities can be used to obtain true prevalence values in the study population.

Moreover, our results indicated a prevalence of mild CHD close to that reported in a past screening study, and the 3 most common CHDs that we observed were VSD, ASD, and PDA. Overall, between 1970 and 2017, the prevalence of CHD increased by 10% every 5 years globally, with over 90% of this increase likely due to the increased detection of relatively mild lesions (VSD, ASD, and PDA).[16] The sex-related CHD patterns identified in our study are similar to those reported in the literature.[27-29] Females are more likely to experience mild CHD. Most ASD cases are found in adulthood. Patent ductus arteriosus is often associated with a family history, which may be related to economic conditions and birth conditions. Congenital heart disease—pulmonary arterial hypertension is common in older men, and most of the cases are associated with extracardiac diseases.

Extracardiac abnormalities have been reported in approximately 20% of CHD cases. These findings are consistent with those of other studies.[30,31]

Although our screening reflects the prevalence of CHD in 518 258 people and represents a diverse population within China, it is still not a population epidemiologic study. At present, the level of CHD diagnosis in China has reached that in developed countries. Many coastal provinces have established CHD screening and treatment networks. However, due to

disparities in economic and medical resources, many children with CHD in deprived areas in Yunnan Province are still not diagnosed and treated in a timely manner. We recommend developing AI methods to screen and diagnose CHD and improve the diagnosis rate of CHD in these areas. With the implementation of these recommendations, morbidity and mortality due to CHD can be reduced at the national level.

The heart sounds included in the CHD-HSY database were obtained from several regions and collected in either clinical or nonclinical environments from pathological patients. A total of 1892 heart sound recordings were obtained from 5 locations. Abnormal heart sounds were recorded from patients with a confirmed cardiac diagnosis. The redundant noise was removed from the heart sounds by filtering. The normalized heart sounds were processed using an event-synchronous segmentation procedure. With the above experimental results, we can draw the following conclusions. Congenital heart disease heart sounds are significantly different from normal heart sounds and can be distinguished and classified by pattern recognition and AI technology. CHD-PAH can be distinguished and classified by pattern recognition and AI technology. At present, the number of CHD heart sound samples is limited. Increasing the number of heart sound samples could enhance the verification of the existing pattern recognition algorithms based on AI, improve these algorithms, and improve the recognition rate of various diseases. The limited data size constrains the development of state-of-the-art deep learning techniques. However, more advanced data augmentation methods, such as generative adversarial networks,[32,33] which were recently successfully applied for snore sound recognition, could be developed.[34] Unlike typical audio signals, e.g., speech or music signals, heart sounds are physiological signals. Therefore, more advanced signal processing methods should be considered.

For instance, wavelet transformation was found to be very efficient in our studies.[35-39] Annotating heart sound data is an expensive and time-consuming task that requires professionally trained experts in cardiology. To reduce future work in expert annotation, active learning,[40,41] and cooperative learning[42] could be introduced and applied to these data. These types of fundamental knowledge, such as the relationship between acoustic representations and anatomical changes in the heart, should be investigated in detail in future work.

Study Limitation

The limitation of our study is that the study has a small number of patients. Further prospective studies with a larger number of patients are needed on this issue. Another limitation is not all the people involved in the screening have been confirmed by echocardiography. Only the suspected people were examined by echocardiography, which may lead to some patients not being found.

CONCLUSION

Artificial intelligence-based auxiliary diagnosis achieves high accuracy in CHD screening and diagnosis. This approach is an effective auxiliary screening method that is worth popularizing in areas without high-level auscultation or ultrasonic Doppler available.

Ethics Committee Approval: Ethical committee approval was received from the Ethics Committee of Yunnan University, (Approval No: CHSRE2021008).

Informed Consent: Written informed consent was waived for this study given its retrospective design.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception: Hongbo Yang and Tao Guo; Design: Hongbo Yang and Tao Guo; Supervision: Tao Guo; Data collection and/or processing: Hongbo Yang, Jiahua Pan and Tengyuan Ma; Analysis and/or interpretation: Weilian Wang; Writer: Hongbo Yang; Critical review: Hongbo Yang, Jiahua Pan, Weilian Wang, Tao Guo and Tengyuan Ma.

Acknowledgments: The authors would like to thank the researchers at Yunnan University, who collected and annotated the heart sound corpus. The CHD-HSY corpus (train and dev sets) will be released partially to the public for research use in future challenges or workshops.

Declaration of Interests: The authors declare no conflicts of interest.

Funding: This work was funded by the Major Science and Technology Projects of Yunnan Province (grant no. 2018ZF017-105), the National Natural Science Foundation of China (grant no. 81960067), and the Yunnan Applied Basic Research Project (grant no. 2018FE001). Thanks to Fuwai Yunnan Cardiovascular Hospital for providing a clinical research environment and medical guidance for this study.

REFERENCES

1. Blue GM, Kirk EP, Sholler GF, Harvey RP, Winlaw DS. Congenital heart disease: current knowledge about causes and inheritance. *Med J Aust*. 2012;197(3):155-159. [CrossRef]
2. Liu Y, Chen S, Zühlke L, et al. Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol*. 2019;48(2): 455-463. [CrossRef]
3. Dixit R, Rai SK, Yadav AK, et al. Epidemiology of congenital heart disease in India. *Congenit Heart Dis*. 2015;10(5):437-446. [CrossRef]
4. van der Linde D, Konings EE, Slager MA, et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2011;58(21):2241-2247. [CrossRef]
5. Hoffman JIe. The global burden of congenital heart disease. *Cardiovasc J Afr*. 2013;24(4):141-145. [CrossRef]
6. Lara DA, Lopez KN. Public health research in congenital heart disease. *Congenit Heart Dis*. 2014;9(6):549-558. [CrossRef]
7. Dolk H, Loane M, Garne E. The prevalence of congenital anomalies in Europe. *Adv Exp Med Biol*. 2010;686:349-364. [CrossRef]
8. Ministry of Health of the People's Republic of China. *The Chinese National Report on Birth Defects*. Available at: http://www.gov.cn/gzdt/att/att/site1/20120912/1c6f6506c7f811bacf9301.pdf.
9. Zhang YF, Zeng XL, Zhao EF, Lu HW. Diagnostic value of fetal echocardiography for congenital heart disease: A systematic review and meta-analysis. *Med (Baltim)*. 2015;94(42):e1759. [CrossRef]

10. Abid D, Elloumi A, Abid L, et al. Congenital heart disease in 37,294 births in Tunisia: birth prevalence and mortality rate. *Cardiol Young*. 2014;24(5):866-871. [CrossRef]
11. Roy D, Sargeant J, Gray J, Hoyt B, Allen M, Fleming M. Helping family physicians improve their cardiac auscultation skills with an interactive CD-ROM. *J Contin Educ Health Prof*. 2002;22(3): 152-159. [CrossRef]
12. Mangione S. Cardiac auscultatory skills of physicians-in-traini ng: a comparison of three English-speaking countries. *Am J Med*. 2001;110(3):210-216. [CrossRef]
13. Franklin RC, Jacobs JP, Krogmann ON, et al. Nomenclature for congenital and paediatric cardiac disease: historical perspectives and the International Pediatric and Congenital Cardiac Code. *Cardiol Young*. 2008;18(suppl 2):70-80. [CrossRef]
14. Ismail S, Siddiqi I, Akram U. Localization and classification of heart beats in phonocardiography signals—a comprehensive review. *EURASIP J Adv Signal Process*. 2018;2018(1):26. [CrossRef]
15. Syed Z, Leeds D, Curtis D, Nesta F, Levine RA, Guttag J. A framework for the analysis of acoustical cardiac signals. *IEEE Trans Bio Med Eng*. 2007;54(4):651-662. [CrossRef]
16. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. National time trends in congenital heart defects, Denmark, 1977-2005. *Am Heart J*. 2009;157(3):467-473.e1. [CrossRef]
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. [CrossRef]
18. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag*. 2012;29(6):82-97. [CrossRef]
19. Krizhevsky A, Sutskever I, Hinton GE. *Imagenet classification with deep convolutional neural networks*. Paper presented at: Proceedings of the NIPS, Stateline, NV. 2017;60(6):1097-1105. [CrossRef]
20. Szegedy C, Toshev A, Erhan D. *Deep Neural Networks for Object Detection*. Paper presented at: Proceedings of the NIPS, Stateline. NV; 2013:2553-2561.
21. Cummins N, Baird A, Schuller BW. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. *Methods*. 2018;151:41-54. [CrossRef]
22. Elman JL. Finding structure in time. *Cogn Sci*. 1990;14(2):179-211. [CrossRef]
23. Dai L, Zhu J, Liang J, Wang YP, Wang H, Mao M. Birth defects surveillance in China. *World J Pediatr*. 2011;7(4):302-310. [CrossRef]
24. Ho TC, Ouyang H, Lu Y, Young AH, Chintala K, Detrano RC. Post-procedural outcomes of rural children undergoing correction of congenital heart lesions in Yunnan Province, China. *Pediatr Cardiol*. 2011;32(6):811-814. [CrossRef]
25. Wu L, Li B, Xia J, et al. Prevalence of congenital heart defect in Guangdong Province, 2008-2012. *BMC Public Health*. 2014;14: 152. [CrossRef]
26. Zhang X, Li S, Wu S, et al. Prevalence of birth defects and risk-factor analysis from a population-based survey in Inner Mongolia, China. *BMC Pediatr*. 2012;12:125. [CrossRef]
27. Forrester MB, Merz RD. Descriptive epidemiology of selected congenital heart defects, Hawaii, 1986-1999. *Paediatr Perinat Epidemiol*. 2004;18(6):415-424. [CrossRef]
28. Reller MD, Strickland MJ, Riehle-Colarusso T, Mahle WT, Correa A. Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. *J Pediatr*. 2008;153(6):807-813. [CrossRef]
29. Lindinger A, Schwedler G, Hense HW. Prevalence of congenital heart defects in newborns in Germany: results of the first registration year of the PAN Study (July 2006 to June 2007). *Klin Padiatr*. 2010;222(5):321-326. [CrossRef]
30. Massin MM, Astadicko I, Dessy H. Noncardiac comorbidities of congenital heart disease in children. *Acta Paediatr*. 2007;96(5): 753-755. [CrossRef]
31. Øyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. Recurrence of congenital heart defects in families. *Circulation*. 2009;120(4):295-301. [CrossRef]
32. Schuller B, Steidl S, Batliner A, et al. *The INTERSPEECH 2018 Computational Paralinguistics Challenge: Atypical & Self-Assessed Affect, Crying & Heart Beats*. Paper presented at: Proceedings of the INTERSPEECH, Hyderabad, India: ISCA; 2018:122-126. [CrossRef]
33. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D. *Generative adversarial networks*. Paper presented at: Proceedings of the NIPS, Montreal, Canada. 2020:2672-2680.
34. Zhang Z, Han J, Qian K, Janott C, Guo Y, Schuller B. Snore-GANs: improving automatic snore sound classification with synthesized data. *IEEE J Biomed Health Inform*. 2020;24(1):300-310. [CrossRef]
35. Qian K, Janott C, Zhang Z, Heiser C, Schuller B. *Wavelet features for classification of vote snore sounds*. Paper presented at: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP); Shanghai, China; 2016:221-225. [CrossRef]
36. Qian K, Janott C, Deng J, et al. *Snore sound recognition: on wavelets and classifiers from deep nets to kernels*. Paper presented at: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC); Jeju Island, Korea; 2017:3737-3740. [CrossRef]
37. Qian K, Janott C, Pandit V, et al. Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis. *IEEE Trans Bio Med Eng*. 2017;64(8):1731-1741. [CrossRef]
38. Qian K, Janott C, Zhang Z, et al. Teaching machines on snoring: a benchmark on computer audition for snore sound excitation localisation. *Arch Acoust*. 2018;43(3):465-475.
39. Qian K, Schmitt M, Janott C, et al. A bag of wavelet features for snore sound classification. *Ann Biomed Eng*. 2019;47(4): 1000-1011. [CrossRef]
40. Qian K, Zhang Z, Baird A, Schuller B. Active learning for bird sounds classification. *Acta Acust U Acust*. 2017;103(3):361-364. [CrossRef]
41. Qian K, Zhang Z, Baird A, Schuller B. Active learning for bird sound classification via a kernel-based extreme learning machine. *J Acoust Soc Am*. 2017;142(4):1796. [CrossRef]
42. Zhang Z, Coutinho E, Deng J, Schuller B. Cooperative learning and its application to emotion recognition from speech. *IEEE/ACM Trans Audio Speech Lang Process*. Paper presented at: IEEE/ACM. 2014:1-1. [CrossRef]