Research on Intelligent Cardiopulmonary Sound Monitoring System Based on Wavelet Decomposition and Deep Learning

Kairui Liu
School of Electrical Engineering
Beijing Jiaotong University
Beijing, China
22291009@bjtu.edu.cn

Shihang Wang
School of Electrical Engineering
Beijing Jiaotong University
Beijing, China
22291248@bjtu.edu.cn

Fuyao Ban*
School of Electrical Engineering
Beijing Jiaotong University
Beijing, China
22291227@bjtu.edu.cn

Zixiao Wang
School of Future
Beijing University of Posts and Telecommunications
Beijing, China
ntb666@bupt.edu.cn

Yangjia Hu
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
huyangjia@mail.ustc.edu.cn

Heze Liu
School of Electrical Engineering
Beijing Jiaotong University
Beijing, China
23121445@bjtu.edu.cn

***Abstract*—Addressing the shortcomings of traditional stethoscopes and existing electronic stethoscopes in terms of objective analysis, intelligence, and continuous monitoring, this paper proposes a smart cardiopulmonary sound monitoring system based on piezoelectric sensors, wavelet decomposition, and deep learning. The system hardware achieves signal acquisition and wireless transmission through high-sensitivity piezoelectric sensors and an embedded microcontroller. On the software level, it employs Butterworth filtering combined with wavelet decomposition techniques for efficient noise reduction processing. Utilizing the PyTorch framework, intelligent classification of normal/abnormal cardiopulmonary sound states is achieved based on the ResNeXt deep learning model, whose performance is superior to various existing models. The system integrates signal acquisition, processing, monitoring, and alarming functionalities, providing a feasible solution for the convenient, accurate, and continuous monitoring and auxiliary diagnosis of cardiopulmonary sounds.**

Keywords—electronic stethoscope, piezoelectric sensor, wavelet decomposition, deep learning, cardiopulmonary sounds, intelligent monitoring

I. Introduction

According to the World Health Organization (WHO) report [1], the world's leading causes of death (ranked by total fatalities) are primarily associated with two categories: cardiovascular diseases and respiratory system diseases. Specifically, ischemic heart disease is the largest cause of death globally, accounting for 13% of total fatalities, while chronic obstructive pulmonary disease (COPD) ranks as the third leading cause, responsible for 5% of total deaths.

Clinically, cardiopulmonary sounds are two highly important physiological signals of the human body. They serve as crucial evidence for disease diagnosis and are a primary means for physicians to assess cardiovascular and respiratory system conditions. As the primary tool for physicians' auscultation, the performance quality of the stethoscope directly impacts the accuracy of disease diagnosis [2].

Traditional acoustic stethoscopes face significant limitations in diagnosis, relying heavily on physician's experience and auditory capabilities and lacking quantitative objective criteria. Furthermore, the sounds they acquire are transient and cannot be recorded, replayed, or quantitatively analyzed, which hinders disease tracking and remote consultation. Simultaneously, without noise reduction capabilities, they struggle in noisy environments or when identifying subtle abnormal sounds.

In contrast, current mainstream electronic stethoscopes still exhibit various deficiencies in terms of portability and intelligence. They are often limited to basic digitalization and single function, lacking effective intelligent auxiliary diagnosis, consequently offering little significant advantage over traditional stethoscopes. Moreover, they do not provide capabilities for continuous monitoring, making frequent hospital visits for follow-up examinations burdensome for patients, particularly those with chronic illnesses.

To address these shortcomings, the development of a portable electronic stethoscope capable of continuous monitoring and auxiliary diagnosis is highly desirable. Such a device would enable real-time monitoring of patients' health status. Coupled with remote consultation, this could significantly reduce the burden of medical visits for patients and facilitate the timely detection of lesions without requiring hospitalization.

Therefore, building upon the foundations of traditional stethoscopes and existing mainstream electronic stethoscopes, this paper proposes a full-process intelligent cardiopulmonary sound acquisition system. This system covers the entire process from signal acquisition and noise reduction processing to intelligent diagnosis. It not only enables real-time auscultation but also provides functions such as visualization, storage, replay, continuous monitoring, and auxiliary diagnosis, aiming to meet the requirements of continuous monitoring and telemedicine.

II. Hardware Design of the Cardiopulmonary Sound Acquisition System

A. Design Scheme of the Cardiopulmonary Sound Acquisition System

In order to meet the requirements for portability, real-time auscultation, visualization, storage, replay, continuous monitoring, and auxiliary diagnosis, as well as to provide high-precision, low-noise output, the hardware system has been designed with the following key modules: a piezoelectric thin film sensor, an amplification module, a main control module, a power module, and a Bluetooth

979-8-3315-0371-0/25/$31.00 ©2025 IEEE

communication module. The overall functional block diagram of the system is shown in Fig. 1.

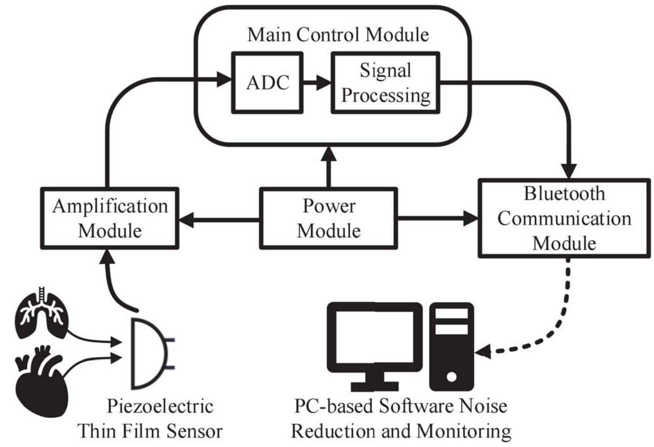

Fig. 1. Functional block diagram of the cardiopulmonary sound acquisition system.

B. Piezoelectric Film Sensor

This system utilizes the CM-01B contact sensor. It is composed of a high-sensitivity, stable piezoelectric film combined with a low-noise preamplifier circuit. Compared to condenser microphones, it offers advantages such as a relatively simple structure, less susceptibility to failure during long-term use, primary sensitivity to contact vibrations, and relative insensitivity to ambient noise, making it more suitable for continuous monitoring application scenarios. The frequency response curve of this sensor is smooth within the 8Hz to 2.2 kHz frequency band, with no significant attenuation [3]. While the frequency range of heart sound signals is 20~150 Hz [4], and that of lung sound signals is 50~2500 Hz [5]. Therefore, this sensor essentially meets the requirements of this system. Fig. 2 shows the frequency response curve of the CM-01B sensor.

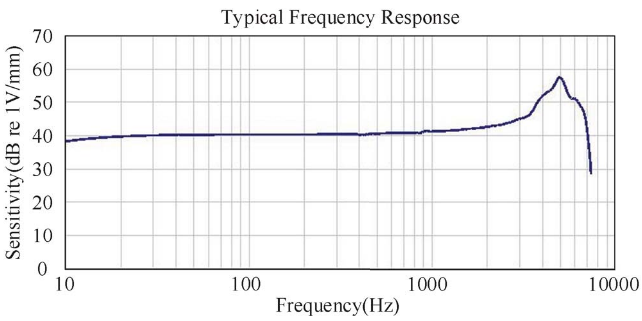

Fig. 2. Frequency response of the CM-01B.

C. Power Amplifier Module

The output signal of the CM-01B piezoelectric film sensor is usually weak. Therefore, an amplification module is needed to perform amplification and noise reduction processing on the signal, to ensure the signal is clear and reliable for subsequent signal processing and display. This is mainly achieved through the following three steps:

(1) Signal Amplification: Amplify the weak voltage signal output from the sensor to a usable amplitude through an amplification circuit.

(2) Noise Suppression: Reduce interference from external ambient noise and internal circuit noise through a low-noise design.

(3) Signal Conditioning: Perform filtering and buffering on the amplified signal to ensure signal quality.

D. Main Control Module

This module primarily contains the MCU (Microcontroller Unit) and its surrounding circuitry. Its main functions are to convert the analog signal from the amplification module into a digital signal and then package it into frames. The processor used is the STM32F103C8T6, and the AD conversion utilizes the integrated 12-bit successive approximation ADC of the STM32F103C8T6. The STM32F103C8T6 offers advantages of low cost, low power consumption, and stability. The input voltage range of the 12-bit ADC is 0 to 3.3 V, and the maximum sampling rate can reach up to 1 MSPS, essentially meeting the requirements of this system. The sampling rate is set to 5 kHz. The 12-bit data obtained from ADC sampling is packaged into frames, with each frame being 5 bytes in size, and transmitted to the PC via the subsequent Bluetooth communication module.

E. Bluetooth Communication Module

The Bluetooth module used in this system is the HC05 [6]. It employs the Bluetooth V2.0 protocol and operates in the 2.4 GHz frequency band. The maximum baud rate is up to 1382.4 kbps. The HC05 Bluetooth module communicates with the STM32 via a serial port. Data is transmitted to the subsequent software system by pairing with the PC.

III. PC End Signal Processing and Auxiliary Diagnosis

The software part of this system utilizes the signals collected by the hardware through a combination of software signal processing and continuous monitoring. The software signal processing part is programmed using MATLAB, which implements the functions of dual-mode noise reduction, visualization, and audio storage. The continuous monitoring part employs deep learning to perform real-time binary classification recognition on the audio files obtained from the signal processing part, while also incorporating functions for re-listening and abnormal state alarming. A GUI (Graphical User Interface) is developed for both parts to improve ease of use.

A. Software Noise Reducing

In the software noise reduction part, this system employs a noise removal method that combines a Butterworth filter with wavelet decomposition noise reduction. Firstly, targeting the distinct frequency bands of heart and lung sounds, two separate Butterworth filters are designed to attenuate high-frequency and low-frequency noise present in the signal. However, traditional band-pass filters struggle to completely remove residual noise within the passband and some aliased signals. To further enhance signal purity, the system integrates a wavelet decomposition algorithm after the filtering stage, performing noise separation on the filtered signal. Through the synergistic application of the Butterworth filter and wavelet decomposition, the system simultaneously ensures the integrity of the main signal frequency bands and significantly enhances the noise reduction effect, thereby achieving high-quality extraction of heart and lung sound signals.

1) Butterworth filtering: Relevant research indicates that the main frequency bands for various pathological heart and lung sounds differ. The primary frequency band for heart sounds is 20-250 Hz, and for lung sounds it is 100-1000 Hz [7]. Therefore, two separate band-pass filters can be designed to remove noise and simultaneously reduce the overlap between the two types of signals.

Comparing the amplitude-frequency and phase-frequency characteristics of Butterworth filters with those of Chebyshev and Bessel filters, it can be intuitively seen that Butterworth filters possess a very flat amplitude response within the passband. Furthermore, their phase characteristic is relatively smooth, which significantly minimizes phase distortion [8]. For these reasons, Butterworth filters are selected for heart and lung sound signal processing. Fig. 3 presents a comparison of the magnitude-frequency characteristic curves and phase-frequency characteristic curves of the three filter types.

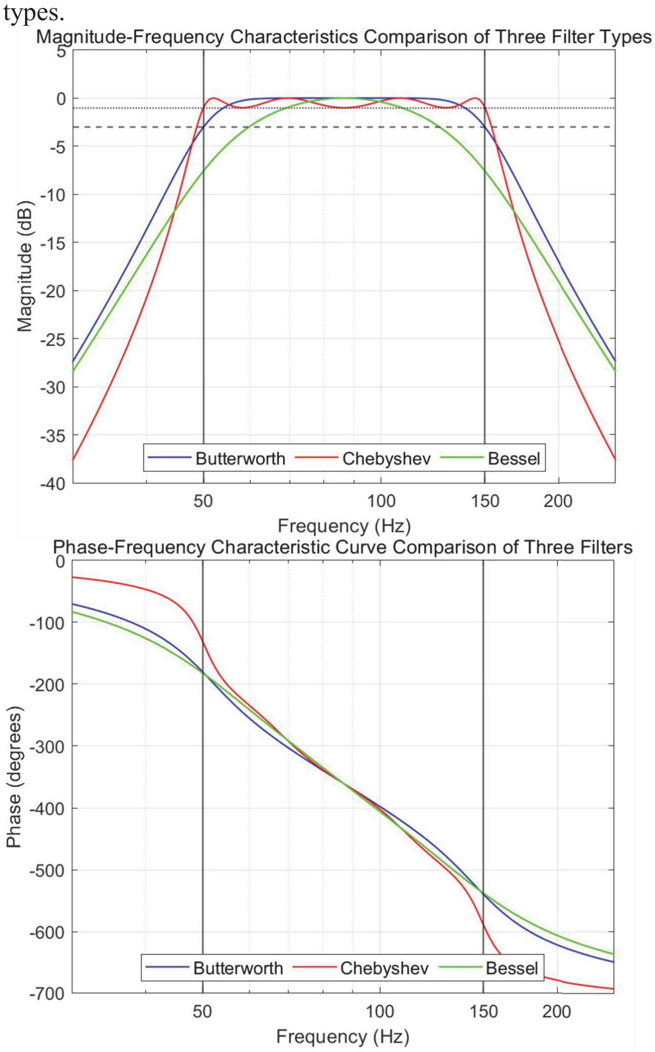

Fig. 3. Bode diagram of three filter types.

2) Wavelet Denoising: Heart and lung sound signals are non-stationary. Traditional time-frequency analysis methods, such as the Short-Time Fourier Transform (STFT), are limited by fixed time-frequency windows, making it difficult to simultaneously achieve high time resolution and high frequency resolution. In contrast, the wavelet transform offers superior time-frequency localization capabilities, adaptively adjusting the time-frequency window. The multi-resolution analysis characteristic of wavelet transform is very suitable for analyzing non-stationary signals, and it can characterize the local features of heart and lung sound signals in the time-frequency domain. This makes wavelet transform widely used in heart and lung sound analysis [9-11].

In this system, after the signal passes through the Butterworth filter, the output data will inevitably contain some noise within the passband. At the same time, the collected heart and lung sound segments also exhibit overlap. By performing wavelet transform analysis and processing on heart and lung sound signals, residual circuit and environmental noise can be suppressed to some extent. Furthermore, by using different wavelet decomposition methods for heart sounds and lung sounds separately and applying targeted denoising to the two signals, the effect of characterizing the features of the two signals separately and reducing overlap can be achieved.

This system uses Daubechies wavelets to process heart and lung sound signals [12]. An N=6 Daubechies wavelet is used for heart sounds, and an N=4 Daubechies wavelet is used for lung sounds. Both undergo 1-level decomposition, balancing computational efficiency and denoising requirements. Using the wavedec function, the signal is decomposed into approximation coefficients cA and detail coefficients cD. Subsequently, the detail components of each level are extracted using detcoef.

Addressing the differences in signal characteristics between heart and lung sounds, a differentiated thresholding strategy is employed. For heart sounds, a global hard threshold (using the wthresh function) is applied to directly filter out low-amplitude noise:

$$cD_{\text{new}} = \begin{cases} cD - \text{thr}, & cD > \text{thr} \\ 0, & |cD| \le \text{thr} \\ cD + \text{thr}, & cD < \text{-thr} \end{cases} \tag{1}$$

where cD represents the detail coefficients and thr is the threshold, which is determined by the statistical properties of the noise in the heart sound frequency band, with priority given to ensuring the amplitude integrity of the S1/S2 heart sounds.

For lung sounds, on the other hand, dynamic threshold calculation is employed, based on the median absolute deviation (MAD) of the detail coefficients for each level to estimate the noise level:

$$\sigma = \text{median}(|cD|)/0.6745 \tag{2}$$

$$\text{thr} = \alpha \cdot \sigma \sqrt{2\log(N)} \tag{3}$$

where α is the layered decay factor (0.3-0.5), which increases with the decomposition level to adapt to the high-frequency noise distribution.

The thresholding employs a continuous threshold function:

$$cD_{\text{new}} = \text{sign}(cD) \cdot \left(|cD| - \frac{\text{thr}}{1 + e^{-|cD|/\text{thr}}} \right) \tag{4}$$

This function smoothly transitions near the threshold, preserving weak lung sound features.

B. Design of Auxiliary Diagnosis Algorithms

In this system, an audio classification method is employed to perform binary classification diagnosis and alarming on the processed cardiopulmonary sounds. The principle is to utilize a deep learning model to classify the processed audio into normal and abnormal categories, and to trigger an alarm for audio classified as abnormal [13].

Audio classification refers to the process of classifying audio signals based on the categories of their content. The purpose is to efficiently organize, retrieve, and understand large volumes of audio data by means of automatic classification. Based on the PyTorch framework, this system utilizes ResNeXt series models, which were trained on a GitHub open-source heart sound dataset and the ICBHI2017 lung sound database [14]. The trained models are subsequently used for intelligent cardiopulmonary sound diagnosis within this system. Fig. 4 presents the process flow diagram of the auxiliary diagnosis system.

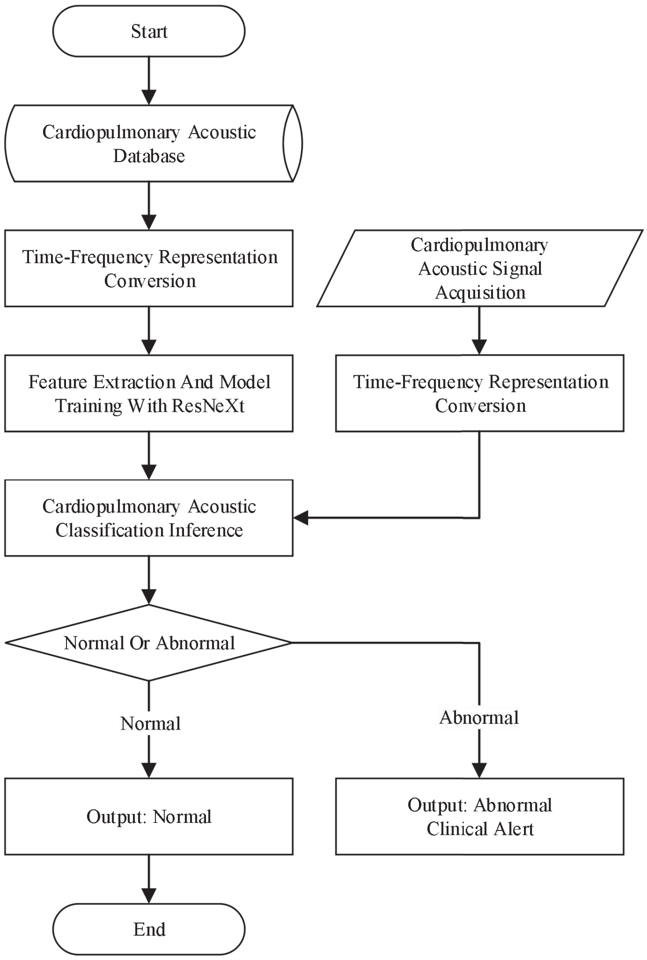

Fig. 4. The process flow diagram of the auxiliary diagnosis system.

This system utilizes PyTorch's highly efficient audio processing interface to convert the collected cardiopulmonary sound signals into Mel spectrograms [15] for subsequent feature extraction and classification. To enhance the richness of feature representation, the Mel spectrogram size is adjusted to 512×512, effectively preserving more detailed information and further enhancing the model's learning capability and recognition performance in high-resolution cardiopulmonary sound classification tasks. Subsequently, the ResNeXt101 network is used to perform deep feature extraction and classification on the spectrograms, enabling the automatic classification of cardiopulmonary sounds.

The ResNeXt101_32x8d model employed by this system is an improved convolutional neural network architecture belonging to the ResNeXt series of models. It introduces the concepts of grouped convolutions and cardinality based on ResNet, significantly enhancing the model's feature representation capability while maintaining efficiency in complex computations. This model has a convolutional neural network depth of 101 layers, a cardinality of 32, and 8 channels per group. The number of layers represents the model depth, while cardinality and the number of channels per group control the model's parallelism. This model improves performance by increasing network width rather than depth, while avoiding parameter explosion. The model output is a binary classification result, classifying the cardiopulmonary sound signals as "normal" or "abnormal", thereby enabling the monitoring and alarming for real-time collected abnormal cardiopulmonary sounds.

At the start of the training process, a warmup strategy is implemented to gradually increase the learning rate from a minimum value of 0 to the set initial learning rate of 0.001 during the initial stage. Additionally, a step-based learning rate decay strategy is added, where the learning rate is reduced to 0.5 times its value every preset period of 50 epochs. This is done to balance the model's convergence speed and accuracy, thereby preventing parameters from oscillating around the optimal value.

Simultaneously, the following five data augmentation strategies [16] are employed: time masking, frequency masking, Gaussian noise, random loudness, and adversarial sample augmentation. This effectively improves the model's generalization capability. Specifically, the time masking strategy involves randomly masking consecutive frames on the time axis of the spectrogram, simulating the absence of audio segments; the frequency masking strategy involves randomly masking consecutive frequency bands on the frequency axis of the spectrogram, simulating the loss of frequency information; the Gaussian noise strategy involves adding random Gaussian noise to the audio waveform, simulating environmental noise interference; the random loudness strategy involves randomly adjusting the audio loudness, simulating different volume scenarios; and adversarial sample augmentation involves generating adversarial perturbations and adding them to the audio, enhancing the model's robustness against adversarial attacks.

IV. System Test Results

During the actual testing process, the sensor's sensing point was placed on the corresponding auscultation site of the test subject, in both heart sound and lung sound modes separately [17]. After adjusting the corresponding parameters on the PC, in both modes, the original signal and the cardiopulmonary sound waveforms after Butterworth filtering and wavelet decomposition could be clearly observed on the GUI interface of the MATLAB program.

By default, the acquisition of cardiopulmonary sound signals is performed automatically. Both types of signals (heart and lung sounds) are acquired every 10 seconds and then processed immediately. During this processing, the waveforms on the GUI are updated, and the audio files are stored simultaneously. Subsequently, the processed signals are delivered to the auxiliary diagnostic system for further analysis.

The heart sound test results are shown in Fig. 5, and the lung sound test results are shown in Fig. 6.

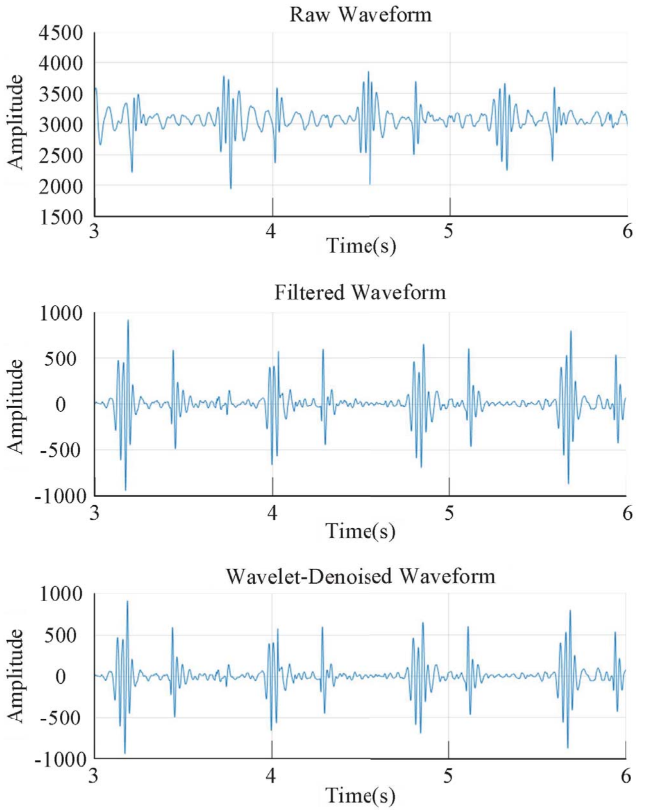

Fig. 5. Heart Sound Test Waveform.

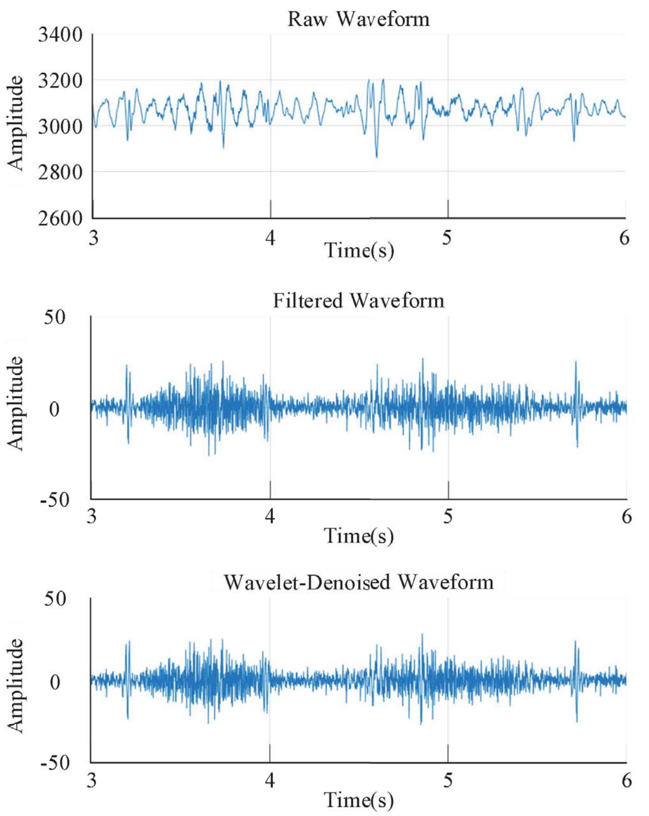

Fig. 6. Lung sound test waveforms.

During the classification validation on the heart sound and lung sound test sets, the proposed ResNeXt101_32x8d model achieves a classification accuracy of 93.52%. To further verify its effectiveness in cardiopulmonary sound diagnosis, this paper compared this model with commonly used cardiopulmonary sound classification models. The results are shown in TABLE I.

TABLE I. COMPARISON OF THE PROPOSED MODEL WITH OTHER COMMON CARDIOPULMONARY SOUND MODELS

Model	Accuracy /%	Training Parameters
AlexNet	75.92	14589636
VGG16	86.68	23569348
GoogLeNet	87.61	10315580
ResNet50	84.13	23516228
GoogLeNet-LSTM	92.06	23107388
ResneXt101_32x8d	93.52	88250280

Notably, the ResneXt101_32x8d model achieves an accuracy of 93.52% on the lung sound test set. This result is more favorable than the diagnostic results achieved by models such as ResNet50 and GoogLeNet-LSTM presented in TABLE I, enabling it to provide users with accurate real-time cardiopulmonary sound diagnosis. After enabling single-channel diagnosis for both heart and lung sounds, the system was designed with a dual-channel mode switching capability. A GUI interface was developed to facilitate switching between the heart sound diagnosis mode and the lung sound diagnosis mode.

Furthermore, the auxiliary diagnosis program can also perform functions such as continuous diagnosis and alarming in both modes respectively. The running results of the auxiliary diagnosis program are shown in Fig. 7.

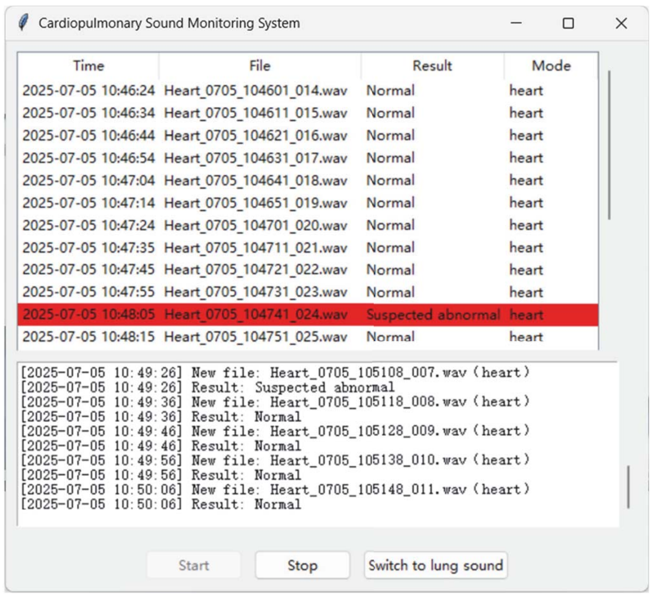

Fig. 7. Auxiliary Diagnosis Program Results.

In terms of real-time performance, with the default signal acquisition duration set to 10 seconds (adjustable via the GUI interface), the signal processing time is approximately 50 milliseconds, typically within 100 milliseconds, and the ResNeXt model takes about 70 milliseconds for classification. Thus, the total processing latency is approximately 120 milliseconds.

Clinical data indicate that the rescue of conditions such as cardiac arrest and sudden cardiac death is extremely urgent. Take cardiac arrest as an example: as one of the most acute-onset and fatal cardiovascular emergencies, it causes most patients to begin suffering irreversible brain damage within 4–6 minutes, followed by a transition to biological death within a few minutes. The critical golden rescue time limit is generally defined as within 4 minutes. Prior to its onset, there is a rapid deterioration in cardiac pumping function, accompanied by characteristic changes in heart sound signs (significant weakening or disappearance of heart sounds, or the occurrence of pathological murmurs/rhythm disorders).

This system supports continuous monitoring. In the presence of pathological signals, it can quickly identify abnormalities and trigger alarms, thereby gaining valuable time for pre-sudden death interventions. The system performs diagnosis every 10 seconds, with the total response cycle accounting for only about 4% of the emergency time window. The processing cycle is far shorter than the progression time of the condition while being longer than the cycle scale of cardiopulmonary sounds. Even if signal fluctuations occur during monitoring, the system can complete analysis and trigger alarms within milliseconds, securing several minutes of rescue time for patients. Therefore, it meets the clinical requirements for real-time performance.

V. Conclusion

This paper developed a smart cardiopulmonary sound monitoring system based on wavelet decomposition and deep learning. By selecting high-sensitivity sensors, employing a software noise reduction method combining Butterworth filtering and wavelet decomposition, and utilizing ResNeXt series models for recognition, the system achieves a complete closed-loop process from signal acquisition and noise reduction processing to intelligent diagnosis. It also includes features such as storage, real-time listening, real-time display, continuous monitoring, and abnormal alarming. This system provides raw data support for subsequent cloud data synchronization, remote consultation, and other related applications, demonstrating broad application prospects.

References

[1] “The top 10 causes of death,” World Health Organization, August 2024. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[2] M. E. Tavel, "Cardiac auscultation: a glorious past—and it does have a future!", Circulation, no. 9, 2006.

[3] H. Y. Li, X. Y. Pan, and K. Yu, “Research on cardiopulmonary sound acquisition system using piezoelectric sensor,” Application of Single-chip Microcomputer and Embedded System, vol. 19, no. 3, pp. 62–66, 2019.

[4] H. Ren, H. Jin, C. Chen, H. Ghayvat, and W. Chen, "A novel cardiac auscultation monitoring system based on wireless sensing for healthcare," IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1–12, 2018.

[5] A. Gurung, C. G. Scrafford, and J. M. Tielsch, "Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis," Respir Med, vol. 105, no. 9, pp. 1396–1403, 2011.

[6] G. P. Sathvik, K. V. Monish, and S. Madhu, "Bluetooth powered real-time monitoring using wireless stethoscope," 2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India, pp. 1–6, 2025.

[7] D. F. Luo, H. Y. Wu, and W. J. Lu, ″Research and development of portable cardiopulmonary sound acquisition equipment,″ (in Chinese), China Medical Device Information, vol. 24, no. 5, pp. 28–31, 2018.

[8] S. Basu, and S. Mamud, "Comparative study on the effect of order and cut off frequency of Butterworth low pass filter for removal of noise in ECG signal," in 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), 2020.

[9] P. K. Jain, and A. K. Tiwari, "An adaptive thresholding method for the wavelet based denoising of phonocardiogram signal," Biomedical Signal Processing and Control, vol. 38, pp. 388–399, 2017.

[10] A. Haar, "On the Theory of Orthogonal Functional Systems," Mathematische Annalen, vol. 69, pp. 331–371, 1910.

[11] M. Ramamonjisoa, M. Firman, and J. Watson, "Single image depth prediction with wavelet decomposition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11089–11098, 2021.

[12] S. N. Ali, S. B. Shuvo, M. I. S. Al-Manzo, A. Hasan, and T. Hasan, "An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise," IEEE Access, vol. 11, pp. 87887–87901, 2023.

[13] T. H. Chen, L. Q. Han, and H. T. Tang, “Research on phonocardiogram signal analysis methods and their applications,” Journal of Beijing Technology and Business University (Natural Science Edition), vol. 27, no. 2, pp. 35–39, 2009.

[14] B. M. Rocha, D. Filos, and L. Mendes, "An open access database for the evaluation of respiratory sound classification algorithms," Physiological Measurement, vol. 40, no. 3, p. 035001, 2019.

[15] N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using local binary pattern," ArXiv preprint arXiv:1710.01703,2017.

[16] T. Xiang, SH. W. Guo; X. G. Li, "Perceptual visual security index based on edge and texture similarities," IEEE transactions on information forensics and security, vol. 11, no. 5, pp. 951–963, 2016.

[17] Y. Torabi, S. Shirani, and J. P. Reilly, "Descriptor: heart and lung sounds dataset recorded from a clinical manikin using digital stethoscope (HLS-CMDS)," IEEE Data Descriptio, 2025.