

## Highlights

### **Rapid auscultation techniques for acute heart failure in ambulance scenarios**

Hui Yu,Zhaoyu Qiu,Zhigang Li,Jinglai Sun,Guangpu Wang,Jing Zhao,Shuo Wang

- An auscultation dataset has been established, containing 2999 recordings from heart failure patients, each with rich annotations, and is publicly accessible on <https://github.com/qiuzhaoyu/AHF-Rapid-Diagnosis>.
- A DenseHF-Net model for diagnosing acute heart failure has been proposed, featuring only 0.33M parameters, making it suitable for rapid auscultation in ambulance scenarios.
- A multi-region fusion auscultation strategy with an accuracy of 99.25% has been proposed, capable of comprehensively analyzing heart sounds from the mitral valve, pulmonary valve, and aortic valve.
- A mitral valve auscultation strategy has been proposed, achieving an accuracy of 92.60% based on 15 seconds of auscultation of the mitral valve.

# Rapid auscultation techniques for acute heart failure in ambulance scenarios

Hui Yu<sup>a</sup>, Zhaoyu Qiu<sup>a</sup>, Zhigang Li<sup>b</sup>, Jinglai Sun<sup>a</sup>, Guangpu Wang<sup>a</sup>, Jing Zhao<sup>a,\*</sup> and Shuo Wang<sup>a,\*</sup>

<sup>a</sup>Department of Biomedical Engineering, Tianjin University, Tianjin, 300072, China

<sup>b</sup>Department of Emergency Medicine, Tianjin 4TH Centre Hospital, Tianjin, 300142, China

## ARTICLE INFO

### Keywords:

Heart Sound Signal  
Acute Heart Failure  
Mel Frequency Cepstrum Coefficient  
Lightweight Deep Learning Model

## ABSTRACT

**Background and Objectives:** Acute Heart Failure (AHF) results in over 26 million hospital admissions worldwide annually, posing a significant global health concern. Currently, AHF diagnosis relies on biochemical markers and echocardiography, which take more than 20 minutes. Auscultation, a quick and non-invasive clinical practice, is used alongside the gold standard. To address the need for rapid clinical AHF diagnosis, this paper presents a model for feature extraction and diagnosis using short heart sound signals.

**Methods:** Discrete wavelet transform was applied for heart sound denoising, and Mel Frequency Cepstrum Coefficients were used for feature extraction. A new DenseHF-Net is proposed for diagnosing heart failure. Additionally, a feature fusion method is introduced for multi-region fusion auscultation, including mitral, aortic, and pulmonic valves, along with an ensemble method for long auscultation in the mitral valve region.

**Results:** An auscultation dataset containing 2999 recordings, each with rich annotations, has been established. The proposed wavelet denoising algorithm achieved a signal-to-noise ratio of 7.8 dB. For multi-region fusion auscultation using DenseHF-Net, the average accuracy reached 99.25%. For mitral valve ensemble auscultation, the average accuracy was 92.60%.

**Conclusions:** The proposed method enables rapid auscultation for AHF, providing accurate results based on a 3-second auscultation recording. While multi-region fusion auscultation achieves high accuracy, mitral valve ensemble auscultation offers a good balance between efficiency and accuracy. This research has the potential to be applied in cardiac auscultation tools for mobile phones, cloud-based systems, or electronic gloves. Data and codes are accessible on <https://github.com/qiuzaoyu/AHF-Rapid-Diagnosis>.

## 1. Introduction

### 1.1. Background

Cardiovascular disease (CVD), including heart failure (HF) and stroke, has become the leading cause of death and disability worldwide [1–4]. The rising age-standardized rate of CVD is especially evident in almost all non-high-income countries [2]. Heart failure, particularly in its acute form (AHF), represents a critical stage in CVD, characterized by the heart's impaired ability to contract and stretch. AHF is a leading cause of emergency hospital admissions, especially among the elderly population [5, 6]. It accounts for over 26 million hospital admissions annually, with a mortality rate ranging from 20–30% [7]. In the management of AHF, emergency procedures such as ambulance response time and door-to-balloon time are crucial, as reducing these intervals can significantly decrease patient mortality and morbidity [8, 9]. For instance, in myocardial infarction—a common cause of AHF—reducing ambulance response time by 15.3 minutes and door-to-balloon time by 36 minutes has been shown to decrease hospital stays by 6.3 days, lower mortality by 12.33%, and reduce rehospitalization rates by 19.69% [9]. Consequently, optimizing these procedures, particularly by shortening door-to-balloon time and conducting rapid

diagnostics such as auscultation during ambulance transport, can greatly enhance survival rates in AHF patients. However, traditional diagnostic methods for AHF, including auscultation, require further optimization. While auscultation is a key rapid diagnostic tool, its effectiveness in ambulance scenarios requires additional quantitative evaluation to ensure it meets the urgent demands of modern emergency care.

Traditional AHF diagnosis relies heavily on clinical biochemical trials, and there is a significant lack of convenient and accurate diagnostic methods, particularly for the 20% of patients with no history of chronic heart failure (CHF). According to the European Society of Cardiology's AHF First Aid Guidelines, the diagnostic criteria for AHF encompass clinical evaluation, electrocardiogram (ECG), chest x-ray, imaging techniques, laboratory tests, and echocardiography [10]. Diagnosis typically hinges on the patient's history and physical examination, supplemented by physical tests such as ECG, echocardiography, and biochemical markers like brain natriuretic peptide (BNP) and NT-proBNP. However, current methods are time-consuming; for example, mainstream NT-proBNP and BNP biochemical detection equipment reports waiting times of 9–16 minutes for BNP results and 11–21 minutes for NT-proBNP results [11]. Additionally, echocardiography, which provides crucial information on the ejection fraction for preoperative evaluation of AHF, requires about 20–30 minutes for a single examination, including approximately five minutes for equipment setup

\*Corresponding author

✉ zhaojing\_zj@tju.edu.cn (J. Zhao); ws111@tju.edu.cn (S. Wang)  
ORCID(s): 0000-0002-7728-7367 (Z. Qiu)

[12]. These time requirements highlight the need for more rapid and accessible diagnostic tools, especially in emergency settings where timely decision-making is critical.

Auscultation, as part of the clinical evaluation, is an effective means of diagnosing heart failure and is recommended by the European Society of Cardiology as a class-1 method [10]. Heart sounds are produced by the mechanical motion of the heart's dynamic system, comprising various mechanical vibrations caused by blood movement within the cardiovascular system. These heart sounds contain a wealth of information regarding the physiology of the cardiovascular system and serve as a crucial auditory representation of the heart's contractile and stretch functions [4, 13, 14]. Normal heart sounds consist of four distinct tones, known as the first heart sound (S1), second heart sound (S2), third heart sound (S3), and fourth heart sound (S4), occurring in sequence during the cardiac cycle. The mitral valve, aortic valve, pulmonary valve, and tricuspid valve are the four most commonly used auscultation areas. The tricuspid valve area, in particular, is crucial for detecting right heart failure. Unlike ECG, which records electrical activity, heart sounds reflect the ventricle's ability to pump blood, providing potentially richer pathological information within a single cardiac cycle. Abnormal heart sounds generally exhibit more background noise, greater fluctuations in S1 amplitude, and more high-frequency details compared to normal heart sounds, making auscultation a valuable tool for early detection of heart dysfunction.

## 1.2. Related works

The current research on digital auscultation technology encompasses aspects such as datasets, signal processing, and diagnostic models.

In terms of datasets, the most commonly utilized dataset is the one established by PhysioNet for the Heart Sound Classification Challenge in 2016 [15], including 665 abnormal heart sounds and 2575 normal heart sounds. Yaseen et al. [16] provided a five-classification dataset of Aortic Stenosis (AS), Mitral Regurgitation (MR), Mitral Stenosis (MS), Mitral Valve Prolapse (MVP) and Normal (N), with 200 cases of each type. Nonetheless, current datasets may have limitations in annotation richness, which can hinder the development of algorithms aimed at diagnosing specific medical conditions. Therefore, to develop rapid diagnostic models for AHF, additional efforts are needed to refine the data and standardize inclusion criteria.

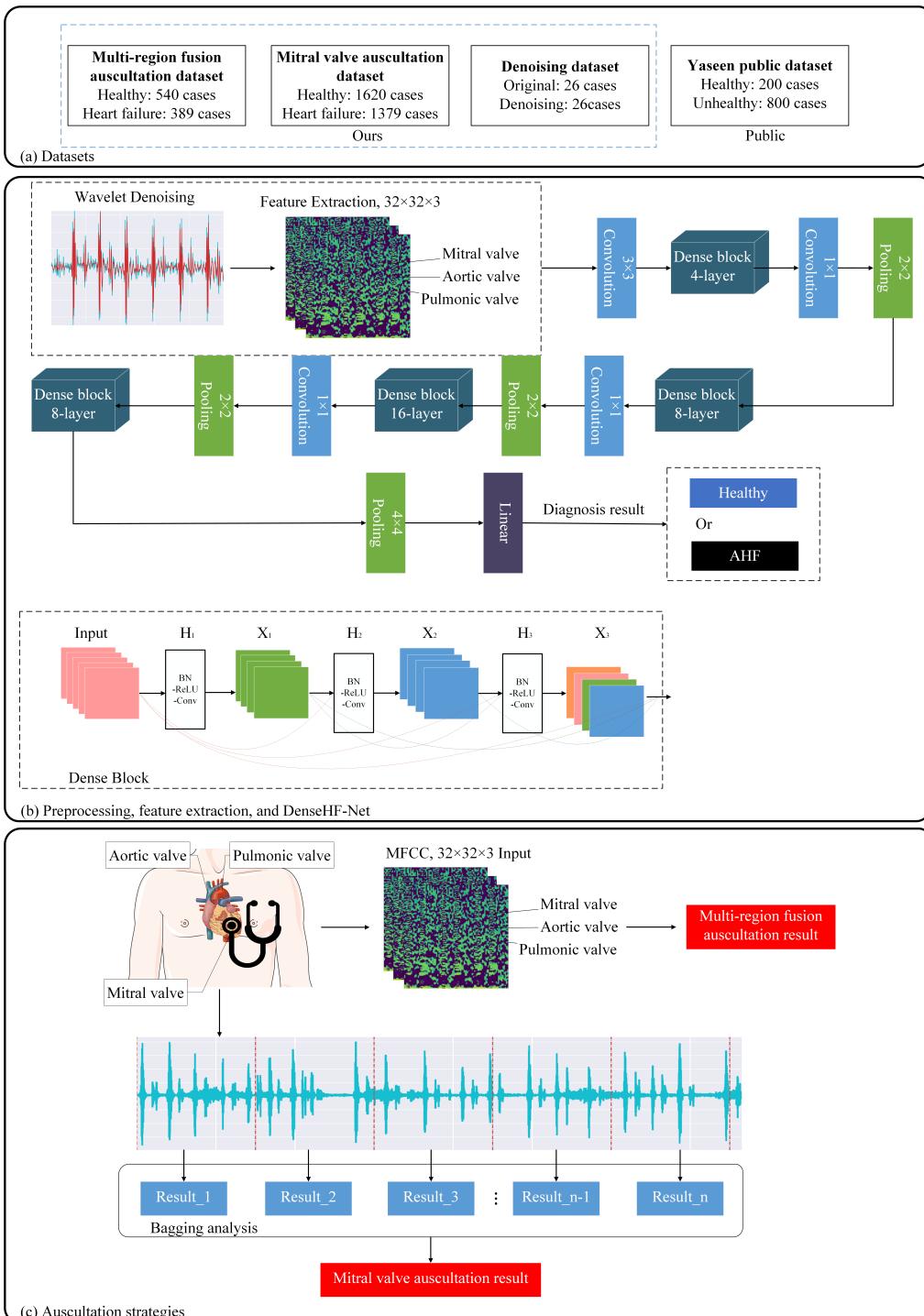
In the realm of pre-processing and feature extraction for heart sounds, various techniques have been developed to analyze these signals effectively. For instance, Vepa's research [17] utilized Short-Time Fourier Transform (STFT) and Discrete Wavelet Transform (DWT) to process heart sound signals. Wu et al. [18] extracted Mel-Frequency Cepstral Coefficients (MFCC) components of heart sound signals based on the hidden Markov model (HMM). Techniques such as MFCC-based feature extraction have proven to be highly successful and widely adopted in heart sound processing. While there are classification algorithms capable of

handling short-duration phonocardiogram (PCG) segments, often as short as 2 seconds with majority voting for final output, many existing methods are optimized for longer signal durations, which may not always be ideal for the rapid diagnosis required in acute heart failure (AHF) scenarios. This highlights the need for further optimization of these techniques to ensure they are suitable for the time-sensitive nature of AHF diagnosis.

In terms of diagnostic models, Rubin [19] used a convolutional neural network (CNN) for heart sound signal classification based on time-frequency characteristics. Arora et al. [20, 21] performed transfer learning of heart sound signals using models like VIZ, MobileNet, Xception, VGG, ResNet, DenseNet, and Inception. Scholars such as Li and Shuvo [22, 23] have developed end-to-end lightweight neural networks for clinical mobile devices. Chen et al. [24] introduced an attention mechanism in the field of heart sound classification and used focal loss as the loss function to address the issue of data imbalance. Existing models are often too large and computationally intensive for the rapid diagnosis of AHF. Therefore, we have developed lightweight models and introduced different auscultation strategies for various clinical scenarios.

To address the challenges of signal processing and diagnostic model in AHF auscultation, our main contributions are:

- An auscultation dataset has been established, containing 2999 recordings from heart failure patients, aiming at addressing the challenges in heart failure auscultation. This dataset encompasses comprehensive information, including diagnostic results, collection area annotations, medical history records, and annotations related to BNP and NT-proBNP. Additionally, this dataset has undergone a rigorous ethical review process and is publicly accessible on <https://github.com/qiuzaoyu/AHF-Rapid-Diagnosis>.
- A wavelet denoising algorithm and a lightweight DenseHF-Net have been developed for short-duration auscultation signals, aiming at rapid diagnosis of acute heart failure. The denoising algorithm proposed in this paper has improved the average signal-to-noise ratio to 7.8 dB. DenseHF-Net has only 0.33M parameters and is easily ported to low-computation scenarios such as mobile terminals.
- We have introduced two auscultation strategies: multi-region fusion auscultation and mitral valve auscultation. Multi-region fusion auscultation is designed for scenarios where long-time auscultation is possible, such as monitoring situations. It utilizes the three most crucial regions, namely, the mitral valve region, aortic valve region, and pulmonic valve region. On the other hand, mitral valve auscultation is specifically designed for rapid diagnosis in cases of heart failure emergencies, focusing solely on the mitral valve auscultation region. It completes the diagnosis within 15 seconds with an accuracy of 92.60%.



**Figure 1: Methodology of the proposed work.** (a) A dataset was built in this paper, including two HF diagnostic subsets, a denoising subset, and a public subset. (b) The layered architecture of the proposed DenseHF-Net contains four Dense blocks. (c) Multi-region fusion auscultation and mitral valve auscultation.

## 2. Materials and Methods

Fig. 1 illustrates the methodology employed in this paper. Firstly, in order to address the data scarcity issue in the development of AHF diagnostic models, we have created several datasets. We create a multi-region fusion auscultation dataset, containing 540 healthy cases and 389

heart failure cases. Each case includes three audio recordings corresponding to the mitral valve, aortic valve, and pulmonic valve. Additionally, a mitral valve auscultation dataset is established, containing 1620 healthy cases and 1379 heart failure cases, with each recording lasting between three to five seconds.

Next, in response to the signal processing and lightweight requirements for rapid AHF diagnosis, we have developed a wavelet denoising algorithm and a lightweight DenseHF-Net. We developed and implemented a wavelet denoising algorithm specifically designed for short-duration heart sound signals, effectively reducing noise. Subsequently, we employ the MFCC algorithm to extract one-dimensional sound signals into two-dimensional image features. Finally, a DenseHF-Net is used to train on the MFCC features.

This paper introduces two distinct auscultation strategies. 1. Multi-region fusion auscultation is designed for scenarios where long-time auscultation is possible, using the fusion characteristics of the mitral valve, aortic valve, and pulmonic valve as input features. 2. Mitral valve auscultation is specifically designed for AHF rapid diagnosis in cases of ambulances. In the case of mitral valve auscultation lasting more than 10 seconds, an ensemble method is proposed.

## 2.1. Datasets

### 2.1.1. HF auscultation datasets

The heart sound databases were acquired at Tianjin 4th Center Hospital of China between 2021 and 2022. The data were recorded using a 3MTM Littmann® electronic stethoscope 3200, with a sampling rate set at 22kHz. This project was approved by the medical ethics committee of Tianjin 4th Center Hospital of China (No. 2022-T050). All volunteers have signed an informed consent.

Under the guidance of two chief physicians, we collected heart sounds from heart failure patients in various departments to create a pathological dataset. We recruited volunteers among patients diagnosed with heart failure, recorded auscultation at three different regions, and simultaneously documented their gender, age, hospitalization information, medical history, and the latest biochemical markers. Afterward, the chief physicians reviewed the data to exclude samples that had already recovered from heart failure or had unclear signs of heart failure features. Finally, heart sounds from a healthy population were collected in the same manner to serve as a comparison group. As shown in Tab. 2, the HF auscultation dataset consists of a total of 71.6 minutes of heart failure auscultation and 81 minutes of the comparison group auscultation.

The multi-region fusion auscultation dataset comprises 540 healthy cases and 389 cases of heart failure. Each case includes three audio recordings of the mitral valve, aortic valve, and pulmonic valve. To ensure robustness, we established a 10-fold cross-validation database after shuffling.

The mitral valve auscultation dataset consists of 1620 healthy cases and 1379 cases of heart failure, with each case featuring audio recordings lasting three to five seconds. Similar to the previous dataset, we created a 10-fold cross-validation database after shuffling.

### 2.1.2. Denoising dataset

As shown in Fig. 1a, we have created a comparative dataset before and after denoising, encompassing 26 different pathological descriptions. The initial 26 recordings are

characterized as noisy signals, encompassing various common abnormal heart sounds. In contrast, the remaining 26 control recordings were subsequently reviewed and verified by two medical professionals to eliminate background noise while preserving all relevant pathological information.

### 2.1.3. Yaseen public dataset

As shown in Fig. 1a, we also use a publicly available Yaseen dataset [16]. This dataset includes Aortic Stenosis (AS), Mitral Regurgitation (MR), Mitral Stenosis (MS), Mitral Valve Prolapse (MVP), and Normal (N). The main purpose is to evaluate the model's generalization ability in diagnosing normal and abnormal heart sounds. We employ 80% training data and 20% testing data. Each case is recorded for a duration of 2 seconds.

## 2.2. Preprocessing

The preprocessing of heart sound signals aims to depress the noisy background in clinical environments. Wavelet transform is used to decrease the noise of heart sounds based on mother wavelets, such as Haar, db, Coif, Sym, and Biorthogonal (bior). Chen et al. [25] achieved optimal denoising results using the db6 wavelet basis. Zhao et al. [26] reported the best outcomes with the bior5.5 basis. Cheng et al. [27] utilized wavelet-based adaptive algorithms to enhance the denoising of heart signals, resulting in a signal improvement of 12.4 dB compared to the pre-denoising state.

In this paper, we consider three wavelet functions: db6, sym8, and coif5. These wavelet bases are used for the discrete wavelet decomposition of heart sound recordings. Following the decomposition, discrete wavelet reconstruction is performed, with a coefficient shrinkage function applied at a threshold of 20% modulo the maximum hard threshold.

Secondly, in order to determine the coefficient contraction strategy, various coefficient contraction functions are applied during both the discrete wavelet decomposition and reconstruction processes.

We introduce a novel self-adaptive threshold function, as depicted in Eq.(1).

$$f_{self}(x, T) = \begin{cases} e^{\frac{x+T}{2}} - e^{\frac{-x-T}{2}} & x \leq -T \\ 0 & -T \leq x \leq T \\ e^{\frac{x-T}{2}} - e^{\frac{-x+T}{2}} & x \geq T \end{cases} \quad (1)$$

$f_{self}$  has the following three advantages:

- $f_{self}$  satisfies:

$$\lim_{x \rightarrow -T^-} f_{self}(x) = \lim_{x \rightarrow T^+} f_{self}(x) = 0$$

$$\lim_{x \rightarrow T^-} f_{self}(x) = \lim_{x \rightarrow T^+} f_{self}(x) = 0$$

$f_{self}$  is differentiable at  $x = \pm T$

- $f_{self}$  is odd, with smooth and monotonically increasing curves.

- $f_{self}$  can overcome the problem of discontinuities in the hard threshold function so that the reconstructed signal retains more detailed information after the reconstruction.

We collect statistical data on the average signal-to-noise ratio (SNR) under different coefficient contraction functions and thresholds.

### 2.3. Feature extraction

Feature extraction for heart sound signals aims to reduce the dimensionality of the data, highlight key information, and thereby improve the subsequent data processing and analysis. Mel Spectrum and MFCC have widely employed feature extraction methods in speech recognition. Human perception of frequency is non-linear, with greater sensitivity to low-frequency signals compared to high-frequency ones. Consequently, frequency conversion is performed according to the equation Eq.(2).

$$Mel(f) = 2595 \ln \left( 1 + \frac{f}{700} \right) \quad (2)$$

To reproduce the Meier scale in the processing of discrete digital signals, the power spectral estimates of the resulting periodic plot are filtered using a Mehr filter bank (usually 26 V-Band Pass filter banks), as shown in Eq.(3).

$$H_m(k) = \begin{cases} \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & others \end{cases} \quad (3)$$

Multiply with FFT to get the Mel spectrum:Eq. (4).

$$MelSpec(m) = \sum_{k=f(m-1)}^{f(m+1)} H_m(k) * |X(k)|^2 \quad (4)$$

Calculate the logarithmic energy output of the V-Band Pass filter bank: Eq.(5). Discrete cosine transform (DCT): Eq. (6) to obtain the MFCC coefficient.

$$S(m) = \ln \left( \sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k) \right), 0 \leq m \leq M \quad (5)$$

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos \left( \frac{\pi n(m-0.5)}{M} \right), n = 1, 2, \dots, L \quad (6)$$

The L order refers to the order of the MFCC coefficient, usually 12-16. M is the number of triangular filters.

The MFCC feature design in this paper is defined as Wu et al. [18] and has achieved the same time-frequency extraction effect as Vepa [17], as shown in Fig. 1a.

### 2.4. DenseHF-Net

The common deep learning model architecture for heart sound diagnosis includes Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and Convolutional Neural Network-Long Short-Term Memory network hybrids (CNN-LSTM) [19, 20, 22, 23]. This paper focuses on model design specifically tailored for AHF rapid diagnosis, aiming to develop a model that balances lightweight characteristics with accuracy.

We introduce DenseHF-Net, which is based on the CVPR 2017 Best Paper, Dense-Net [30]. To achieve model lightweight, DenseHF-Net employs only four Dense Blocks: DenseBlock-1, DenseBlock-2, DenseBlock-3, and DenseBlock-4. This choice reduces the model's parameter count and computational load while still maintaining a certain level of depth and feature extraction capability. Three transition layers with small compression rates are employed simultaneously to reduce the output channel numbers of the 1x1 convolutional layers, thereby reducing the size of feature maps. Ultimately, the diagnostic results are obtained after passing through a linear layer. The parameter settings for the three models are detailed in Tab. 1.

The classic ResNet [28] and MobileNet [29] models are also trained for comparison with DenseHF-Net. All input features are resized to  $32 \times 32$  to make it more convenient for mobile terminals. The parameter numbers of the three models are 0.33M, 42.61M, and 12.24M. The Memory Access Cost (MAC) of the three models are 30.64M, 556.97M, and 46.28M.

Experimental environment system: Ubuntu 18.04; CPU: Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz; GPU: NVIDIA GeForce RTX 3090 with 24G VRAM; CUDA Version: 11.4; Pytorch Version: 1.10.

### 2.5. Auscultation strategies

The multi-region fusion auscultation strategy is designed for AHF diagnosis in hospitals, with the aim of reducing the door-to-balloon time. Then to reduce the mortality rate and rehospitalization rate of HF patients [9]. Heart sound signals are synchronously collected from three regions, including the mitral valve region, aortic valve region, and pulmonic valve region.

The processing pipeline consists of three main steps:

1. Applying wavelet transform to reduce noise in the three heart sound channels.
2. Feature extraction using the MFCC.
3. Fusion of the three MFCC feature sets to produce input of the same dimension as that of the mitral valve auscultation.

The mitral valve auscultation strategy is designed for emergency medical services (EMS) or general screening scenarios, with a stronger emphasis on convenience, aiming to complete the diagnosis within 15 seconds.

As shown in Fig. 1c, for mitral valve auscultation durations exceeding 10 seconds, there arises the need to address the challenge of reducing false positive diagnoses. To tackle

**Table 1**  
Models parameter setting

DenseHF-Net (ours)			ResNet-18 [28]			MobileNetV1-28 [29]		
Number of parameters: 0.33M			Number of parameters: 42.61M			Number of parameters: 12.24M		
Memory access cost: 30.64M			Memory access cost: 556.97M			Memory access cost: 46.28M		
Forward/backward pass size: 20.18M			Forward/backward pass size: 13.63M			Forward/backward pass size: 10.31M		
Layer	Filters	Output	Layer	Filters	Output	Layer	Filters	Output
Feature Map	$[3 \times 3, 24] \times 1$	$32 \times 32$	Feature Map	$[3 \times 3, 64] \times 1$	$32 \times 32$	Feature Map	$[3 \times 3, 32] \times 1$	$30 \times 30$
Dense Block	$[1 \times 1, 48] \times 4$	$32 \times 32$	Conv2_x	$[3 \times 3, 64] \times 2$	$32 \times 32$	Conv_dw,pw	$[3 \times 3, 32] \times 1$	$30 \times 30$
Transition	$[1 \times 1, conv]$ $[2 \times 2, pool]$	$16 \times 16$	Conv3_x	$[3 \times 3, 128] \times 2$	$16 \times 16$	Conv_dw,pw	$[3 \times 3, 64] \times 1$	$15 \times 15$
Dense Block	$[1 \times 1, 48] \times 8$	$16 \times 16$	Conv4_x	$[3 \times 3, 256] \times 2$	$8 \times 8$	Conv_dw,pw	$[3 \times 3, 128] \times 1$	$15 \times 15$
Transition	$[1 \times 1, conv]$ $[2 \times 2, pool]$	$8 \times 8$	Conv5_x	$[3 \times 3, 512] \times 2$	$4 \times 4$	Conv_dw,pw	$[3 \times 3, 128] \times 1$	$8 \times 8$
Dense Block	$[1 \times 1, 48] \times 16$	$8 \times 8$	Pooling	$4 \times 4$	$1 \times 1 \times 512$	Conv_dw,pw	$[3 \times 3, 256] \times 1$	$8 \times 8$
Transition	$[1 \times 1, conv]$ $[2 \times 2, pool]$	$4 \times 4$	Linear		$1 \times 2$	Conv_dw,pw	$[3 \times 3, 256] \times 1$	$4 \times 4$
Dense Block	$[1 \times 1, 48] \times 8$	$4 \times 4$				Conv_dw,pw	$[3 \times 3, 512] \times 5$	$4 \times 4$
Pooling	$4 \times 4$	$1 \times 1 \times 384$				Conv_dw,pw	$[3 \times 3, 512] \times 1$	$2 \times 2$
Linear		$1 \times 2$				Conv_dw,pw	$[3 \times 3, 1024] \times 1$	$2 \times 2$
						Pooling	$2 \times 2$	$1 \times 1 \times 1024$
						Linear		$1 \times 2$

this issue, this research paper introduces an ensemble learning method as a strategic approach.

$$\begin{aligned} Result_i &= \max_{\alpha_i} Softmax(\alpha_i) \\ &= \max_{\alpha_i} \frac{\exp(\alpha_i)}{\sum_{i=1}^n \exp(\alpha_i)} \end{aligned} \quad (7)$$

Eq.7 is used to calculate the diagnostic results for a single fragment.  $\alpha_i$  represents the model output.  $Result_i$  is 0 for healthy and 1 for heart failure.

$$Output = Result_1 \vee Result_2 \vee \dots \vee Result_n \quad (8)$$

Eq.8 represents the result of a one-to-one OR operation applied to auscultation segments, designed to minimize the occurrence of false positives. Here, 'n' denotes the number of fragments, typically set to 3.

### 3. Results

#### 3.1. Datasets

Tab. 2 shows the details of the dataset in this paper. The details include the age and gender composition of auscultation volunteers, and case descriptions for the denoising dataset. Auscultation dataset contains 2999 recordings from heart failure patients, each with rich annotations, and are publicly accessible on <https://github.com/qiuzhaoyu/AHF-Rapid-Diagnosis/tree/main/Database>.

### 3.2. Preprocessing

Firstly, we conduct an analysis to determine the average SNR under various combinations of wavelet bases and decomposition layers, using a coefficient shrinkage function based on the 20% modulo maximum hard threshold. The results of this analysis are presented in Fig. 2, which clearly indicates that the Sym8 base at the 7-layer decomposition level yields the most effective denoising results among the tested methods.

Secondly, we further investigate the average SNR across different shrinkage functions and threshold values while maintaining the sym8 base at the 7-layer decomposition level. Our findings indicate that the use of the 20% modulo maximum with the  $f_{self}$  threshold function emerged as the optimal denoising method.

Finally, the average SNR of the denoising algorithm proposed in this paper is 7.8 dB.

### 3.3. Multi-region fusion auscultation

The quality of models refers to Eq.9: Accuracy (Acc), Eq.10: Sensitivity (Se), Eq.11: Specificity (Sp) and Eq.12: F1-Score.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

$$Se = \frac{TP}{TP + FN} \quad (10)$$

**Table 2**

The details of the dataset information.

HF datasets			
Group	Age( $\bar{x} \pm sd$ )	Sex	Length(min)
Healthy	24	Male	27.0
Healthy	$26 \pm 2$	Female	54.0
HF	$72.6 \pm 12.0$	Male	34.8
HF	$77.9 \pm 11.1$	Female	36.8

Denoising dataset		Length(s)
Description		
Systolic murmur		68
Functional aortic stenosis		4
Musical noise		23
Organic mitral regurgitation		59
Functional mitral stenosis		9
Organic mitral stenosis		9
Diplogue		39
Paradoxically divided		13
Varied S1		23
Weak S1		3
Split S1		4
Strong S1		6
Weak S2		29
Split S2		12
Ventricular septal defect		9
Open flap sound		10
Late contraction		9
Relative mitral regurgitation		4
Sinus bradycardia		7
Sinus tachycardia		4
Functional pulmonary regurgitation		3
Pulmonary stenosis		67
Diastolic tetra tone		15
Continuous murmur		15
Overlapping galloping sound		9
Pendulum sound		6

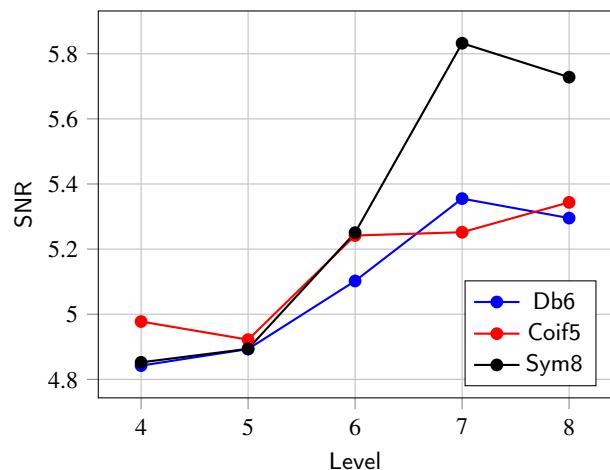
  

Yaseen dataset		Number
Description		
Aortic Stenosis		200
Mitral Regurgitation		200
Mitral Stenosis		200
Mitral Valve Prolapse		200
Normal		200

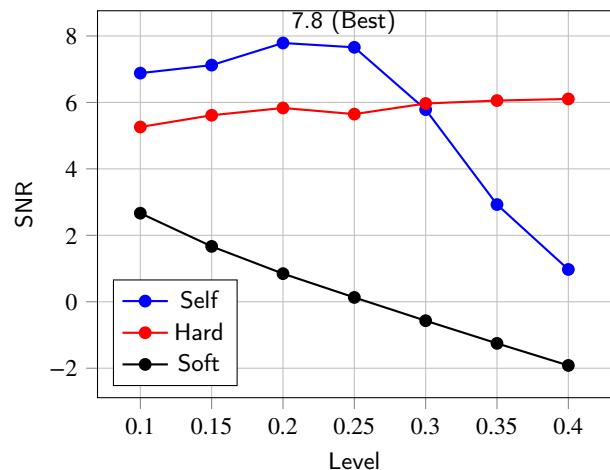
$$Sp = \frac{TN}{TN + FP} \quad (11)$$

$$F1 - Score = 2 \times \frac{Se \times Sp}{Se + Sp} \quad (12)$$

Tab.3 presents the classification results on the HF-Diagnosis dataset and the public Yassen dataset, comparing the performance of three models: DenseHF-Net, ResNet-18, and MobileNetV1-28.



(a) Choice of base and levels



(b) Choice of threshold parameters

**Figure 2: Wavelet denoising of short signal.** (a) Best: wavelet denoising with sym8 base at 7-layer decomposition. (b) Best: wavelet denoising with 20% modulo maximum  $f_{self}$ .

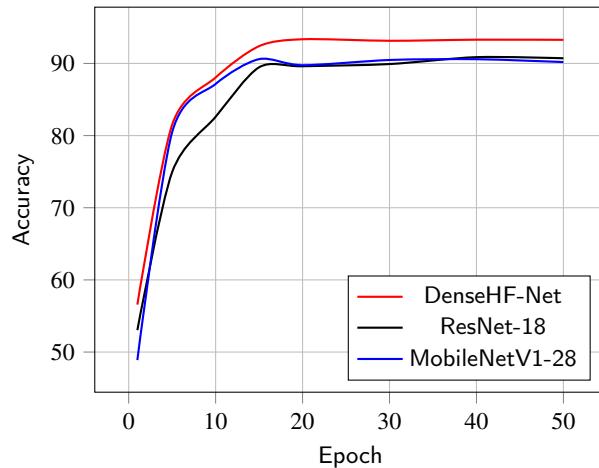
DenseHF-Net achieves the highest accuracy ( $99.25\% \pm 0.71\%$ ) and specificity ( $99.44\% \pm 0.75\%$ ), demonstrating strong classification performance with an F1-score of  $99.10\% \pm 0.64\%$ . In comparison, ResNet-18 has an accuracy of  $98.71\% \pm 1.26\%$  with lower specificity ( $98.10\% \pm 2.68\%$ ), though its sensitivity is close to DenseHF-Net. MobileNetV1-28 excels in sensitivity ( $100.00\% \pm 0.00\%$ ) but has slightly lower specificity ( $97.95\% \pm 1.54\%$ ), with overall performance similar to DenseHF-Net. In summary, DenseHF-Net offers the best balance between sensitivity and specificity, making it the preferred model for HF-Diagnosis tasks.

Fig.3 provides a comprehensive overview of the average performance across the mitral valve auscultation dataset. Notably, ResNet-18 and MobileNetV1-28 exhibit comparable performance, while DenseHF-Net exhibits the most rapid rate of improvement. Importantly, all three models exhibit effective convergence of the loss function. It is

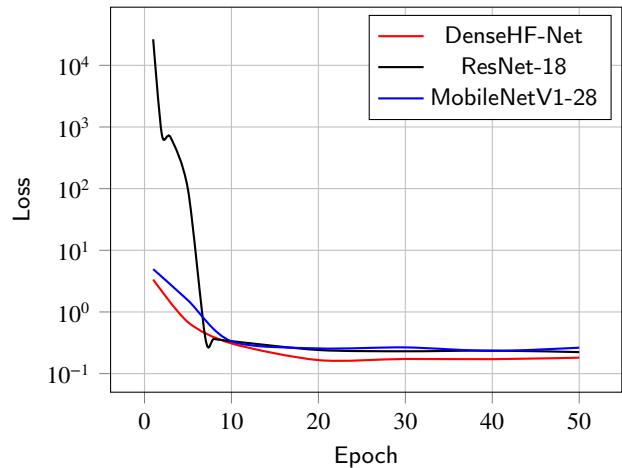
**Table 3**

Classification results of HF-Diagnosis dataset and public Yassen Dataset.

Multi-region fusion auscultation results												
DenseHF-Net(ours)	Average $\pm$ sd	ResNet-18	Average $\pm$ sd									
Acc(%)	99.25 $\pm$ 0.71	Acc(%)	98.71 $\pm$ 1.26									
Se(%)	98.97 $\pm$ 1.22	Se(%)	99.09 $\pm$ 0.92									
Sp(%)	99.44 $\pm$ 0.75	Sp(%)	98.10 $\pm$ 2.68									
F1 – Score(%)	99.10 $\pm$ 0.64	F1 – Score(%)	98.59 $\pm$ 1.49									
Mitral valve auscultation results												
DenseHF-Net(ours)	Average $\pm$ sd	ResNet-18	Average $\pm$ sd									
Acc(%)	92.60 $\pm$ 1.11	Acc(%)	91.33 $\pm$ 2.48									
Se(%)	92.68 $\pm$ 0.88	Se(%)	92.32 $\pm$ 2.59									
Sp(%)	92.53 $\pm$ 2.08	Sp(%)	90.20 $\pm$ 4.49									
F1 – Score(%)	92.01 $\pm$ 2.15	F1 – Score(%)	91.17 $\pm$ 2.58									
Yaseen dataset results												
DenseHF-Net(ours)		ResNet-18										
	AS	MR	MS	MVP	AS	MR	MS	MVP	AS	MR	MS	MVP
Acc(%)	91.25	58.75	75.00	63.75	92.50	88.75	95.00	83.75	82.50	92.50	85.00	88.75
Se(%)	90.00	97.50	75.00	97.50	95.00	80.00	90.00	80.00	87.50	87.50	87.50	87.50
Sp(%)	92.50	20.00	75.00	30.00	90.00	97.50	100.00	87.50	77.50	97.50	82.50	90.00
F1 – Score(%)	91.23	33.19	75.00	45.88	92.43	87.89	94.74	83.58	82.20	92.23	84.93	88.73



(a) Mitral valve auscultation testing accuracy.



(b) Mitral valve auscultation testing loss value.

**Figure 3: Average 10-fold CV history. (a)** Average testing accuracy of three models. **(b)** Average testing loss of three models (without augmentation).

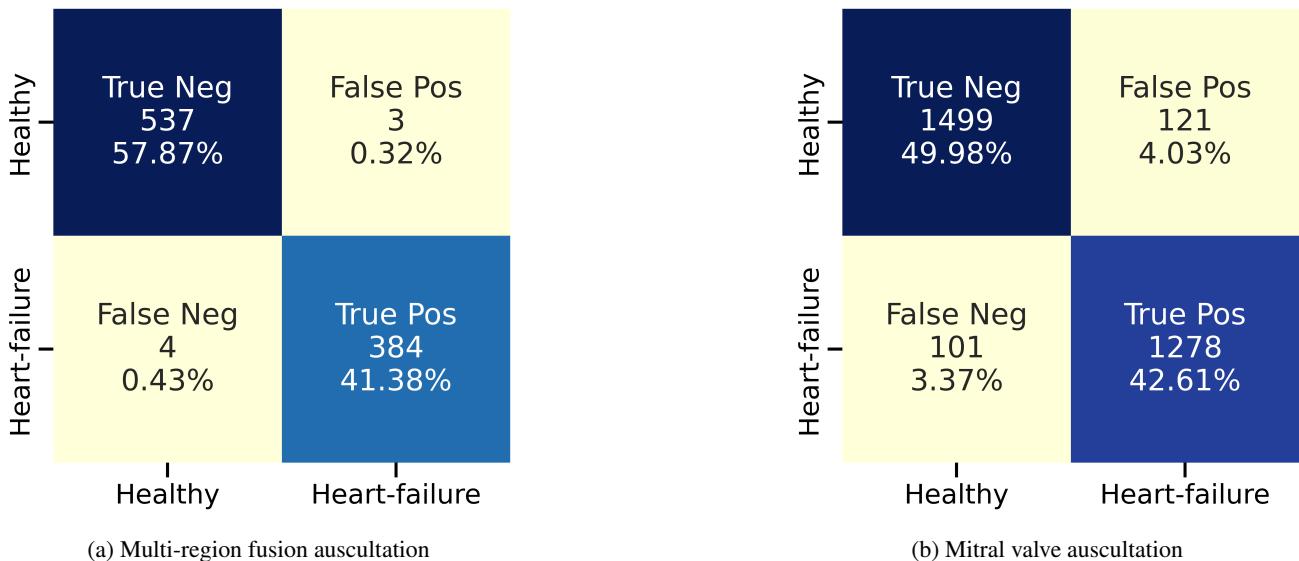
worth highlighting that both ResNet-18 and MobileNetV1-28 demonstrate similar performance trends, with DenseHF-Net demonstrating the fastest convergence among them.

Tab.3 also presents the results obtained from the Yaseen dataset. DenseHF-Net, ResNet-18, and MobileNetV1-28 typically reached convergence around 50 epochs. The AS diagnosis shows the best performance, with sensitivity and specificity exceeding 82% across all three models. For MS and MR, all three models achieve correct diagnoses, with sensitivity and specificity exceeding 87% in both ResNet-18 and MobileNetV1-28. However, MVP diagnosis by DenseHF-Net yields suboptimal results, with a specificity of only 30% and an F1-Score of only 45.88%.

### 3.4. Mitral valve auscultation

For mitral valve auscultation, DenseHF-Net demonstrates the best performance with an accuracy of 92.60%  $\pm$  1.11%, along with a good balance between sensitivity (92.68%  $\pm$  0.88%) and specificity (92.53%  $\pm$  2.08%), as shown in Tab.3. In comparison, ResNet-18 has slightly lower accuracy (91.33%  $\pm$  2.48%), with competitive sensitivity (92.32%  $\pm$  2.59%) but lower specificity (90.20%  $\pm$  4.49%), indicating a higher rate of false positives. MobileNetV1-28 shows the lowest accuracy (90.73%  $\pm$  1.80%) and the least balanced performance in terms of sensitivity (90.68%  $\pm$  2.49%) and specificity (90.80%  $\pm$  2.97%).

Overall, DenseHF-Net is the most effective model for mitral valve auscultation, offering superior accuracy and a

**Figure 4: Comparison of two auscultation strategies**

better balance between sensitivity and specificity, making it the preferred choice among the three models.

Fig. 4 illustrates a comparative analysis of two different auscultation strategies, as represented by the respective confusion matrices. Both strategies focus on the detection accuracy of mitral valve abnormalities. The confusion matrices in Fig. 4b and Fig. 4a summarize the performance outcomes, including True Positives, True Negatives, False Positives, and False Negatives.

In the first strategy (Fig. 4b), the model achieved a high sensitivity of 92.68% and specificity of 92.53%, with a balanced F1 score of 92.01%. The second strategy (Fig. 4a) demonstrated even higher accuracy, achieving a sensitivity of 98.97% and a specificity of 99.44%, leading to an F1 score of 99.10%.

### 3.5. Ablation Study

Table 3 presents the results of the ablation study, using Params, FLOPs, MACs, and ten-fold cross-validation accuracy as evaluation metrics. DenseHF-Net strikes a balance between computational efficiency and performance, with 0.33 million parameters, 61.29 million FLOPs, and 30.64 million MACs, achieving an accuracy of 93.47%.

When wavelet-denoising was not applied, the accuracy dropped to 93.26%, indicating that wavelet-denoising has a certain positive impact on model performance. Similarly, using Mel-spectrum features resulted in a significant decrease in accuracy to 90.77%, highlighting the crucial role of feature selection in maintaining high performance.

Adjusting the compression rate showed a clear trade-off between model complexity and accuracy. As the compression rate increased from 30% to 70%, the number of parameters, FLOPs, and MACs also increased, leading to a slight improvement in accuracy. However, the highest accuracy (95.1%) was achieved without compression, which

also resulted in the highest computational cost, underscoring the need to balance accuracy and efficiency.

Finally, increasing the number of blocks in the network further raised the number of parameters, FLOPs, and MACs. When the block numbers were set to 8, 16, 32, 16, the accuracy reached 94.74%. This indicates that while deeper networks may offer higher accuracy, they also significantly increase the computational burden, which could be a limiting factor in real-time or resource-constrained applications.

In the final configuration, DenseHF-Net employed a 10% compression rate and block numbers of 4, 8, 16, and 8. This configuration was chosen to prune the model as much as possible while maintaining performance, making it more suitable for emergency scenarios such as use in ambulances.

**Table 4**  
Ablation Study Results

Ablation Condition	Params (M)	FLOPs (M)	MACs (M)	Accuracy (%)
<b>DenseHF-Net</b>	0.33	61.29	30.64	92.60
Without wavelet-denoising	0.33	61.29	30.64	92.26
Feature input: Mel-spectrum	0.33	61.29	30.64	90.77
Compression rate: 30%	0.39	67.01	33.50	93.87
Compression rate: 50%	0.47	74.13	37.06	93.91
Compression rate: 70%	0.57	82.42	41.21	94.01
Without compression	0.78	98.84	49.42	95.10
Blocks numbers: 6 12 24 16	0.99	132.64	66.32	94.35
Blocks numbers: 8 16 32 16	1.47	201.76	100.88	94.74

## 4. Discussion

### 4.1. How long should auscultate for?

The optimal input length for heart sounds in AI models is a subject of debate. Cardiac auscultation offers significant predictive value for cardiac diagnosis, characterized by its rapid and cost-effective nature [31]. In the PhysioNet 2016 dataset, which includes 665 abnormal heart sounds and 2575 normal heart sounds, the lengths of recordings range from 3 seconds to 60 seconds. Meanwhile, the Yaseen dataset consists of 1000 abnormal heart sounds and 200 normal heart sounds, all standardized at a length of 2 seconds.

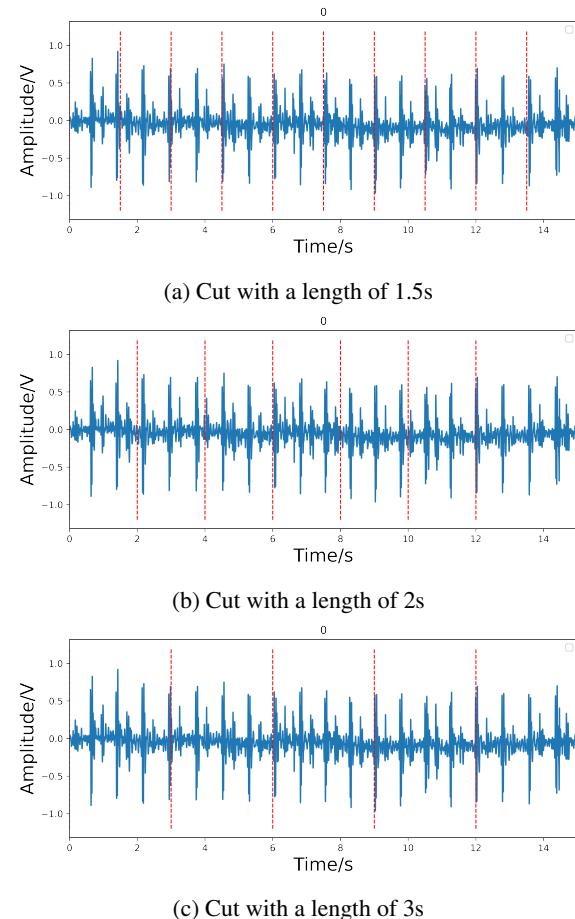
As illustrated in Fig.5, the choice of input length for heart sounds directly impacts the number of cardiac events included, such as S1, diastole, S2, and systole. Fig.5a demonstrates that a 1.5-second input can theoretically encompass two diastolic and two systolic events. Fig.5b shows that a 2-second input can include at least two complete cardiac cycles. Fig.5c reveals that a 3-second input can accommodate three or four full cardiac cycles.

For the purposes of this paper, the input length is standardized at 3 seconds.

### 4.2. Interpretability discussion

Heart sounds contain essential physiological information that reflects the heart's ability to pump blood effectively, and this information is crucial for interpretability in diagnosis. In this study, we analyze the time-frequency characteristics of heart sounds using short-term Fourier transforms (STFT), with a primary focus on the mitral valve, located at the point of the strongest apical beat, to provide a more intuitive explanation.

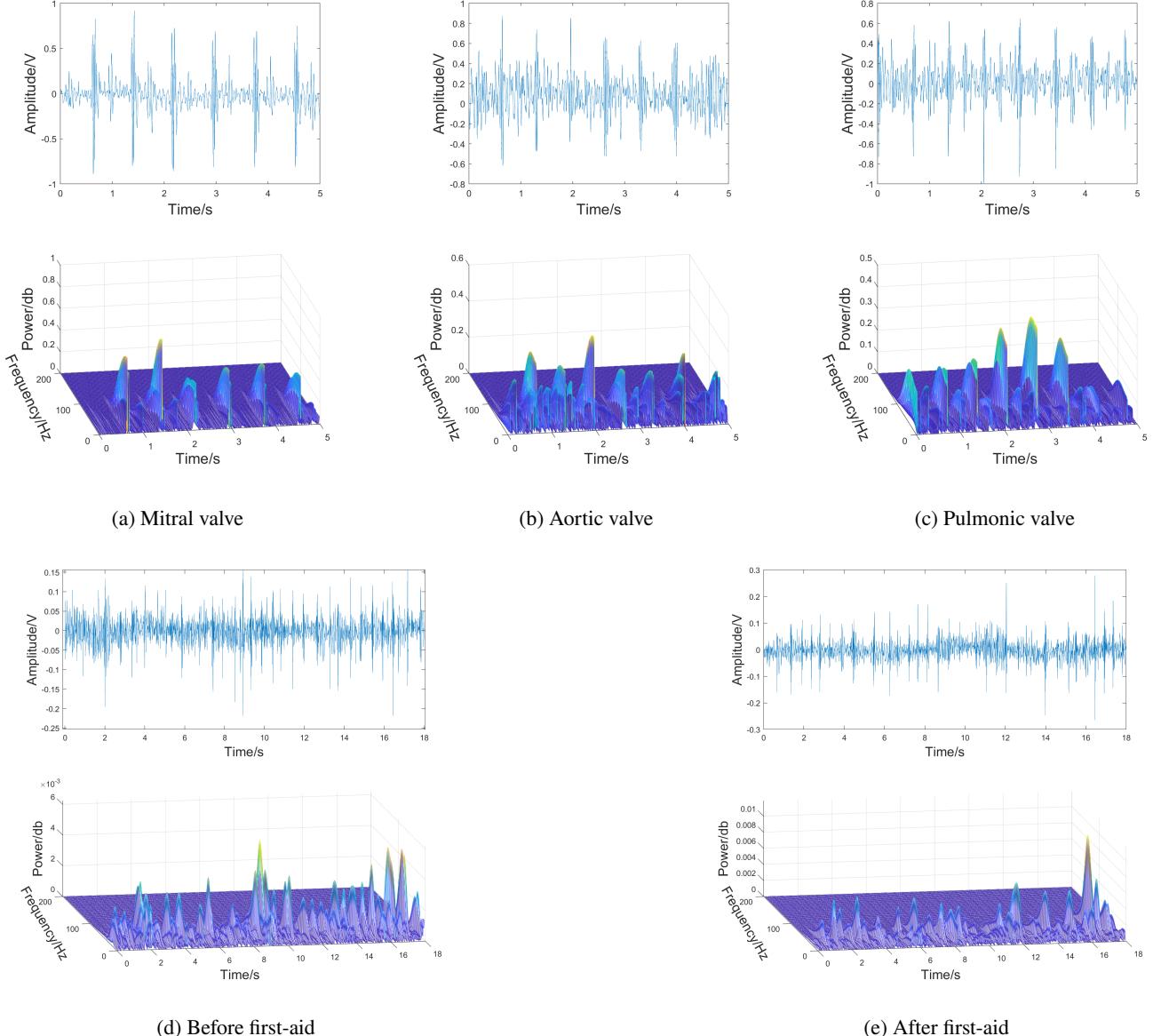
As illustrated in Fig.6a, the stable S1 amplitude is predominantly concentrated within the 100 Hz frequency range, clearly reflecting the normal function of the mitral valve. In contrast, the aortic valve, located in the second intercostal space at the right sternal border, is situated further from the heart and is influenced by lung sounds, resulting in a lower signal amplitude. This variation can be explained by the attenuation of energy during sound propagation. The pulmonic valve, positioned in the second intercostal space at the left sternal border, exhibits signal characteristics that are intermediate between those of the mitral and aortic



**Figure 5: Effect of different cutting lengths.** (a) A heart sound is cut with lengths of 1.5s. (b) The same heart sound is cut with lengths of 2s. (c) The same heart sound is cut with lengths of 3s.

valves, further illustrating the impact of different heart valve locations on heart sound characteristics.

Heart sounds from the mitral valve were collected from the same patient both before and after receiving initial medical intervention, as shown in Fig.6d and Fig.6e, providing interpretability of treatment effects. Before treatment, the S1 amplitudes were unstable and were accompanied by noisy lung sounds, indicating impaired heart function. However,



**Figure 6: Heart sounds in five different situations.** (a) Heart sounds from the mitral valve. (b) Heart sounds from the aortic valve. (c) Heart sounds from the pulmonic valve. (d) Heart sounds before first-aid. (e) Heart sounds after first-aid.

after initial treatment, which took place two days later, a noticeable improvement was observed. The heart sound cycle became clearer, and the energy amplitude of S1 was more pronounced, clearly reflecting the positive impact of the treatment on heart function.

### 4.3. Results discussion

#### 4.3.1. Preprocessing and feature extraction

This paper conducts an in-depth investigation into the optimal decomposition layers, wavelet bases, and threshold shrinkage functions for the denoising of short heart sounds. Fig.2 presents our findings, indicating that a 7-layer decomposition using the  $f_{self}$  threshold based on the sym8 wavelet yields the most effective denoising results. It's worth noting that Db6 and sym8 wavelets exhibit similar properties, but

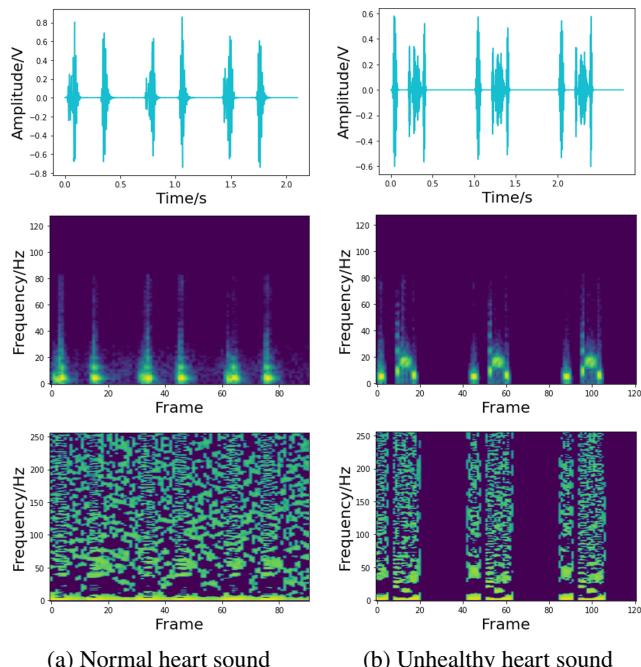
sym8, due to its shorter support length and superior energy concentration, is more morphologically aligned with heart sounds.

In contrast to previous studies by Chen [25] and Zhao [26], this paper employs Signal-to-Noise Ratio (SNR) as the primary metric for evaluating noise reduction, avoiding the limitations of waveform comparison. Additionally, our denoising experiments encompass 26 pathological heart sounds, enhancing the clinical relevance of our findings. Furthermore, in comparison to Cheng [27], we propose a new  $f_{self}$  construction method that significantly improves SNR (7.8 dB).

As depicted in Fig.7, this study delves into the intricacies of feature extraction, particularly focusing on Mel-spectrum and MFCC. The MFCC feature extraction method for heart

sounds possesses the advantage of preserving not only temporal waveform features but also capturing frequency energy distribution. This approach retains valuable information aligned with the Mel scale, which corresponds to human auditory perception.

Fig.7a illustrates a representation of normal heart sounds, while Fig.7b showcases an instance of heart sounds from a patient with Mitral Valve Prolapse. Both Mel spectrum and MFCC representations encompass a comprehensive range of waveform characteristics, effectively capturing the intricacies of high-frequency signal features in the form of heat maps.



**Figure 7: Feature extraction result.** (a) Waveform, Mel-spectrum, MFCC of normal heart sound. (b) Waveform, Mel-spectrum, MFCC of unhealthy MVP heart sound.

#### 4.3.2. Diagnostic performance

Previous researchers have explored various intelligent algorithms for heart failure diagnosis or prediction, as shown in Tab. 5. For instance, Khade [32] and Rao [33] predicted heart failure incidence rates based on physiological parameters or medication records, while Acharya [34] and Matsumoto [35] examined the potential of diagnosing heart failure using ECG or X-ray images. In comparison to these works, our study offers enhanced medical diagnostic value and interpretability, especially in ambulance settings, with superior diagnostic speed, reduced device size, and more suitable application.

Moreover, when compared to previous classification research efforts [17–20, 22, 23], this study incorporates well-established methods such as wavelet denoising, MFCC feature extraction, and CNNs, addressing issues of signal processing and diagnosis in rapid short-duration auscultation.

However, challenges remain in the generalization capabilities for diagnosing mitral regurgitation, with the model still showing signs of overfitting. Additionally, the study lacks a thorough discussion of AHF subtypes.

In comparison, Chen [24] employed Short-Time Fourier Transform (STFT) as a feature extraction method combined with a CNN enhanced by an Attention Mechanism, achieving an accuracy of 94.12% on the PhysioNet 2016 dataset. The introduction of the attention mechanism significantly improved the focus on relevant heart sound signal parts, leading to more accurate classifications. However, Chen's approach, while robust, increases computational costs, making it less suitable for real-time applications in resource-constrained environments like ambulances.

Xiang [21] explored Logmel features with Transfer Learning using pre-trained models such as MobileNet and Xception, achieving an accuracy of 90.23% on the PhysioNet 2016 dataset. Though slightly underperforming in accuracy compared to Chen's method, Xiang's approach offers better generalization and adaptability to different datasets with minimal additional training. The use of lightweight models like MobileNet is particularly promising for deployment in mobile and low-power devices, making it a suitable choice for real-time diagnostic applications.

In summary, while both Chen and Xiang have made significant contributions to heart sound classification, their methods cater to different needs: Chen's approach excels in accuracy through an attention-enhanced CNN, albeit with higher computational intensity, while Xiang's method offers a balance between performance and efficiency, particularly suitable for mobile and real-time applications.

**Table 5**  
Comparison with other researches

Comparison with other heart failure researches				
Authors	Dataset		Aim	Acc(%)
<b>Ours</b>	HF-Diagnosis dataset		HF diagnosis	99.25
Khade [32]	Physiological parameters		CHF prediction	88.30
Rao [33]	Medication records		CHF prediction	93.00
Acharya [34]	ECG		HF diagnosis	98.97
Matsumoto [35]	X-Ray Images		HF diagnosis	82.00

Comparison with other heart sound researches				
Authors	Feature extraction	Classification method	Databases	Acc(%)
<b>Ours</b>	MFCC	DenseHF-Net	Multi region fusion auscultation	99.25
<b>Ours</b>	MFCC	DenseHF-Net	Mitral valve auscultation	92.60
Chen [24]	STFT	CNN with Attention Mechanism	PhysioNet 2016	94.12
Xiang [21]	Logmel	Transfer Learning (MobileNet, Xception)	PhysioNet 2016	90.23
Vepa [17]	STFT, DWT	kNN, MLP, SVM	Normal, systolic, diastolic records	95.20
Wu [18]	MFCC	HMM	325 cycles of 10-type	95.08
Rubin [19]	MFCC	CNN	PhysioNet 2016	95.20
Arora [20]	STFT	CNN	PhysioNet 2016	89.04
Li [22]	STFT	CNN	PhysioNet 2016	85.00
Shuvo [23]	Time-invariant features	CNN	Yaseen database	99.60

## 5. Conclusion

Cardiac auscultation is a crucial clinical method for detecting heart failure and holds significant promise for rapid heart failure detection. This paper discusses signal processing and diagnostic model techniques for the rapid diagnosis of AHF. The average SNR of heart sounds reached 7.8 dB. The models successfully achieve effective feature learning and effective loss convergence. DenseHF-Net stands out as the top-performing model in heart failure diagnosis, achieving an accuracy of 92.60%. Both auscultation strategies can meet the requirements for emergency vehicles and hospital rooms. Future research will focus on collecting multi-center AHF data for model training, aiming to achieve better generalization performance. Simultaneously, we will actively promote clinical testing and assessment to validate its practicality and explore its clinical emergency application value, ultimately aiming to develop a device similar to that in [36] that could obtain FDA approval. Additionally, future research directions will also include expanding the diagnosis to multiple cardiac diseases to further enhance the model's potential in cardiovascular disease detection.

declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 6. Data and code availability

Data collection has been approved by the medical ethics committee of Tianjin 4th Center Hospital of China (No. 2022-T050). Data and codes are available at <https://github.com/qiuzhaoyu/AHF-Rapid-Diagnosis>.

## 7. Funding

This research was supported by The National Natural Science Foundation of China (No. 72174138). The authors

## References

- [1] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, et al., Heart disease and stroke statistics—2020 update: a report from the american heart association, *Circulation* 141 (2020) e139–e596.
- [2] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, E. J. Benjamin, C. P. Benziger, et al., Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study, *Journal of the American College of Cardiology* 76 (2020) 2982–3021.
- [3] G. A. Mensah, G. A. Roth, V. Fuster, The global burden of cardiovascular diseases and risk factors: 2020 and beyond, 2019.
- [4] E. M. Boorsma, J. M. Ter Maaten, K. Damman, W. Dinh, F. Gustafsson, S. Goldsmith, D. Burkhoff, F. Zannad, J. E. Udelson, A. A. Voors, Congestion in heart failure: a contemporary look at physiology, diagnosis and treatment, *Nature Reviews Cardiology* 17 (2020) 641–655.
- [5] L. Sinnenberg, M. M. Givertz, Acute heart failure, *Trends in Cardiovascular Medicine* 30 (2020) 104–112.
- [6] M. Arrigo, M. Jessup, W. Mullens, N. Reza, A. M. Shah, K. Sliwa, A. Mebazaa, Acute heart failure, *Nature Reviews Disease Primers* 6 (2020) 1–15.
- [7] B. Chapman, A. D. DeVore, R. J. Mentz, M. Metra, Clinical profiles in acute heart failure: an urgent need for a new approach, *ESC heart failure* 6 (2019) 464–474.
- [8] S. M. Victor, A. Gnanaraj, S. Vijayakumar, S. Pattabiram, A. S. Mullasari, Door-to-balloon: where do we lose time? single centre experience in india, *indian heart journal* 64 (2012) 582–587.
- [9] Z. Fan, F. Zhang, Effects of an emergency nursing pathway on the complications and clinical prognosis of patients with acute myocardial infarction, *Int J Clin Exp Med* 14 (2021) 661–668.
- [10] M. Nieminen, M. Böhm, M. Cowie, H. Drexler, G. Filippatos, G. Joudeau, Y. Hasin, et al., Task force on acute heart failure. executive summary of the guidelines on the diagnosis and treatment of acute heart failure: the task force on acute heart failure of the european society of cardiology, *Eur Heart J* 26 (2005) 384–416.
- [11] R. A. Lewis, C. Durrington, R. Condliffe, D. G. Kiely, Bnp/nt-probnp in pulmonary arterial hypertension: time for point-of-care testing?, *European Respiratory Review* 29 (2020).
- [12] P. Menon, P. M. Kapoor, M. Choudhury, Echocardiography for left ventricular assist device patients, *Journal of Cardiac Critical Care TSS* 6 (2022) 155–161.
- [13] M. Johnston, S. P. Collins, A. B. Storrow, The third heart sound for diagnosis of acute heart failure, *Current Heart Failure Reports* 4 (2007) 164–169.
- [14] J. Wynne, The clinical meaning of the third heart sound, *The American Journal of Medicine* 111 (2001) 157–158.
- [15] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, R. G. Mark, Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016, in: 2016 Computing in cardiology conference (CinC), IEEE, 2016, pp. 609–612.
- [16] G.-Y. Son, S. Kwon, Classification of heart sound signal using multiple features, *Applied Sciences* 8 (2018) 2344.
- [17] J. Vepa, Classification of heart murmurs using cepstral features and support vector machines, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2009, pp. 2539–2542.
- [18] H. Wu, S. Kim, K. Bae, Hidden markov model with heart sound signals for identification of heart diseases, in: Proceedings of 20th International Congress on Acoustics (ICA), Sydney, Australia, 2010, pp. 23–27.
- [19] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, K. Srirachan, Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients, in: 2016 Computing in cardiology conference (CinC), IEEE, 2016, pp. 813–816.
- [20] V. Arora, K. Verma, R. S. Leekha, K. Lee, C. Choi, T. Gupta, K. Bhatia, Transfer learning model to indicate heart health status using phonocardiogram (2021).
- [21] M. Xiang, J. Zang, J. Wang, H. Wang, C. Zhou, R. Bi, Z. Zhang, C. Xue, Research of heart sound classification using two-dimensional features, *Biomedical Signal Processing and Control* 79 (2023) 104190.
- [22] T. Li, Y. Yin, K. Ma, S. Zhang, M. Liu, Lightweight end-to-end neural network model for automatic heart sound classification, *Information* 12 (2021) 54.
- [23] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, A. Gumaci, Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings, *IEEE Access* 9 (2021) 36955–36967.
- [24] J. Chen, Z. Guo, X. Xu, L.-b. Zhang, Y. Teng, Y. Chen, M. Woźniak, W. Wang, A robust deep learning framework based on spectrograms for heart sound classification, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2023).
- [25] J. Y. Zhao, H. Y. Liu, H. S. Ma, H. D. Zhou, Research of the approach for the fetal heart sound signal's extracting based on coif5 wavelet transform, *Chinese Journal of Biomedical Engineering* (2006).
- [26] T. Chen, L. Han, S. Xing, Research of de-noising method of heart sound signals based on wavelet transform, *Computer Simulation* 27 (2010) 401–405.
- [27] X. Cheng, Z. Zhang, Denoising method of heart sound signals based on self-construct heart sound wavelet, *Aip Advances* 4 (2014) 087108.
- [28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint arXiv:1704.04861* (2017).
- [30] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
- [31] A. J. Taylor, *Learning Cardiac Auscultation: From Essentials to Expert Clinical Interpretation*, Springer, 2015.
- [32] S. Khade, A. Subhedar, K. Choudhary, T. Deshpande, U. Kulkarni, A system to detect heart failure using deep learning techniques, *Int. Res. J. Eng. Technol. (IRJET)* 6 (2019) 384–387.
- [33] S. Rao, Y. Li, R. Ramakrishnan, A. Hassaine, D. Canoy, J. Cleland, T. Lukasiewicz, G. Salimi-Khorshidi, K. Rahimi, An explainable transformer-based deep learning model for the prediction of incident heart failure, *IEEE Journal of Biomedical and Health Informatics* 26 (2022) 3362–3372.
- [34] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, R. S. Tan, Deep convolutional neural network for the automated diagnosis of congestive heart failure using ecg signals, *Applied Intelligence* 49 (2019) 16–27.
- [35] T. Matsumoto, S. Kodera, H. Shinohara, H. Ieki, T. Yamaguchi, Y. Higashikuni, A. Kiyosue, K. Ito, J. Ando, E. Takimoto, et al., Diagnosing heart failure from chest x-ray images using deep learning, *International Heart Journal* 61 (2020) 781–786.
- [36] M. Fudim, M. Mirro, H. M. Cheng, Audicor remote patient monitoring: Fda breakthrough device and technology for heartfailure management, *JACC. Basic to translational science* 7 (2022) 313–315.