Highlights

Rapid diagnosis technology for acute heart failure based on auscultation

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 are publicly accessible on https://github.com/qiuzhaoyu/AHF-Rapid-Diagnosis/Database.

Rapid diagnosis technology for acute heart failure based on auscultation

Hui Yu[a], Zhaoyu Qiu[a], Zhigang Li[b], Jinglai Sun[a], Guangpu Wang[a], Jing Zhao[a,*] and Shuo Wang[a,*]

[a]Department of Biomedical Engineering, Tianjin University, Tianjin, 300072, China
[b]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) leads to over 26 million hospital admissions worldwide annually and is now a major global health concern. Currently, AHF diagnosis relies on biochemical markers and echocardiography, which takes more than 20 minutes. Auscultation, a quick and non-invasive clinical practice, is used alongside the gold standard. Recognizing the need for rapid clinical AHF diagnosis, this paper presents a model for feature extraction and diagnosis using short heart sound signals.

Methods: In this paper, discrete wavelet transform is applied for heart sound denoising, and the Mel Frequency Cepstrum Coefficient is applied for feature extraction. A new DenseHF-Net is proposed for the diagnosis of heart failure. A feature fusion method is proposed for multi-region fusion auscultation, including mitral, aortic, and pulmonic valves. An ensemble method is proposed for long auscultation in the mitral valve region.

Results: An auscultation dataset containing 2999 recordings has been established, each with rich annotations. A proposed wavelet denoising algorithm achieves a signal-to-noise ratio of 7.8 dB. For multi-region fusion auscultation, using DenseHF-Net, the average accuracy is 99.35%. For mitral valve ensemble auscultation, using DenseHF-Net, the average accuracy is 94.41%.

Conclusions: The above method enables rapid auscultation of AHF, providing accurate results based on a 3-second auscultation recording. Multi-region fusion auscultation achieves good auscultation accuracy, but mitral valve ensemble auscultation provides a good balance between efficiency and accuracy. The above research has the potential to be used for cardiac auscultation that can be used on mobile phones, cloud, or electronic gloves.

1. 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 explore a wavelet denoising algorithm specifically designed for short-duration heart sound signals to reduce noise effectively. 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.

Two distinct auscultation strategies are introduced in this paper. 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.

1.1. Datasets

1.1.1. 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 3M™ 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

*Corresponding author
zhaojing_zj@tju.edu.cn (J. Zhao); ws111@tju.edu.cn (S. Wang)
ORCID(s): 0000-0002-7728-7367 (Z. Qiu)

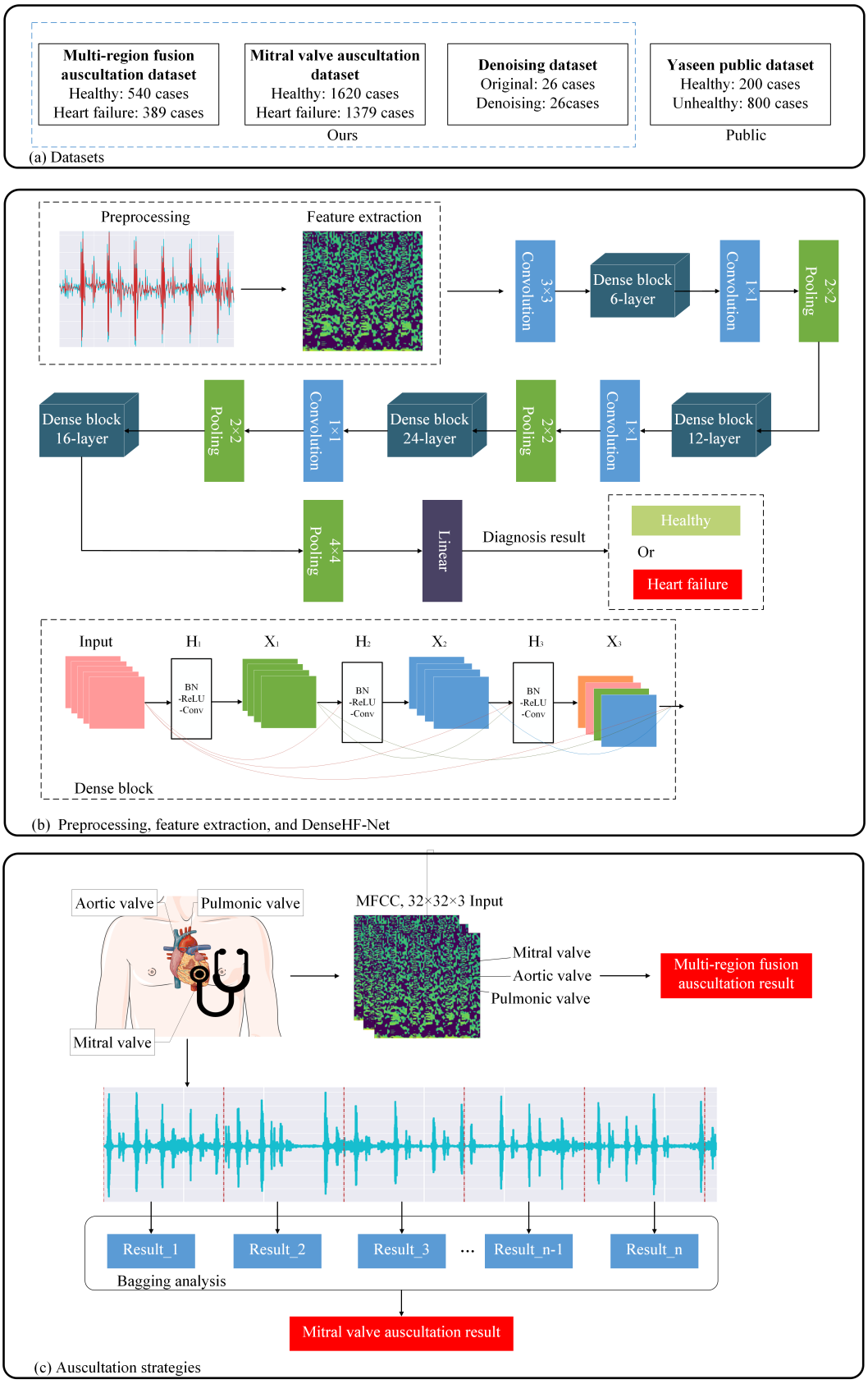


Figure 1: Methodology of the proposed work. (**a**) A dataset was built in this paper, including two HF diagnostic subsets, a denoising subset, and a public subset. (**b**) The layered architecture of the proposed DenseHF-Net contains four Dense blocks. (**c**) Multi-region fusion auscultation and mitral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

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.

1.1.2. 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.

1.1.3. Yaseen public dataset

As shown in Fig. 1a, we also use a publicly available Yaseen dataset [1]. 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.

1.2. 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. [2] achieved optimal denoising results using the db6 wavelet basis. Zhao et al. [3] reported the best outcomes with the bior5.5 basis. Cheng et al. [4] 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).

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

f_{self} has the following three advantages:

- f_{self} satisfies:

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

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

  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.

1.3. 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).

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

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).

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

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

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

Table 1
Models parameter setting

DenseHF-Net (ours)			**ResNet-18** [7]			**MobileNetV1-28** [8]		
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	$\begin{bmatrix}3\times3, 24\end{bmatrix}\times1$	32×32	Feature Map	$\begin{bmatrix}3\times3, 64\end{bmatrix}\times1$	32×32	Feature Map	$\begin{bmatrix}3\times3, 32\end{bmatrix}\times1$	30×30
Dense Block	$\begin{bmatrix}1\times1, 48\\3\times3, 12\end{bmatrix}\times4$	32×32	Conv2_x	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	32×32	Conv_dw,pw	$\begin{bmatrix}3\times3, 32\\1\times1, 64\end{bmatrix}\times1$	30×30
Transition	$\begin{bmatrix}1\times1, conv\\2\times2, pool\end{bmatrix}$	16×16	Conv3_x	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	16×16	Conv_dw,pw	$\begin{bmatrix}3\times3, 64\\1\times1, 128\end{bmatrix}\times1$	15×15
Dense Block	$\begin{bmatrix}1\times1, 48\\3\times3, 12\end{bmatrix}\times8$	16×16	Conv4_x	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	8×8	Conv_dw,pw	$\begin{bmatrix}3\times3, 128\\1\times1, 128\end{bmatrix}\times1$	15×15
Transition	$\begin{bmatrix}1\times1, conv\\2\times2, pool\end{bmatrix}$	8×8	Conv5_x	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	4×4	Conv_dw,pw	$\begin{bmatrix}3\times3, 128\\1\times1, 256\end{bmatrix}\times1$	8×8
Dense Block	$\begin{bmatrix}1\times1, 48\\3\times3, 12\end{bmatrix}\times16$	8×8	Pooling	4×4	$1\times1\times512$	Conv_dw,pw	$\begin{bmatrix}3\times3, 256\\1\times1, 256\end{bmatrix}\times1$	8×8
Transition	$\begin{bmatrix}1\times1, conv\\2\times2, pool\end{bmatrix}$	4×4	Linear		1×2	Conv_dw,pw	$\begin{bmatrix}3\times3, 256\\1\times1, 512\end{bmatrix}\times1$	4×4
Dense Block	$\begin{bmatrix}1\times1, 48\\3\times3, 12\end{bmatrix}\times8$	4×4				Conv_dw,pw	$\begin{bmatrix}3\times3, 512\\1\times1, 512\end{bmatrix}\times5$	4×4
Pooling	4×4	$1\times1\times384$				Conv_dw,pw	$\begin{bmatrix}3\times3, 512\\1\times1, 1024\end{bmatrix}\times1$	2×2
Linear		1×2				Conv_dw,pw	$\begin{bmatrix}3\times3, 1024\\1\times1, 1024\end{bmatrix}\times1$	2×2
						Pooling	2×2	$1\times1\times1024$
						Linear		1×2

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.

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

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

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. [5] and has achieved the same time-frequency extraction effect as Vepa [6], as shown in Fig. 1a.

1.4. 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) [9–12]. 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 [13]. 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 [7] and MobileNet [8] models are also trained for comparison with DenseHF-Net. All input features are resized to 32×32 to make it more convenient for mobile terminals. The parameter numbers of the three models are 3.82M, 42.61M, and 12.24M. The Memory Access Cost (MAC) of the three models are 130.89M, 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.

1.5. 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 [14]. Heart sound

signals are synchronously collected from three regions, including the mitral valve region, aortic valve region, and pulmonic valve region.

The processing pipeline includes three main steps:

1. Applying wavelet transform to the three channels of heart sound to reduce noise. 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 this issue, this research paper introduces an ensemble learning method as a strategic approach.

$$\begin{aligned} Result\_i &= \max_{\alpha_i} Softmax(\alpha_i) \\ &= \max_{\alpha_i} \frac{\exp(\alpha_i)}{\sum_{i=1}^{2} \exp(\alpha_i)} \end{aligned} \tag{7}$$

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

$$Output = Result\_1 \vee Result\_2 \vee \cdots \vee Result\_n \tag{8}$$

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.

2. Results

2.1. 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/jimmytju/AHF-Rapid-Diagnosis/Database.

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 ± 2	Female	54.0
HF	72.6 ± 12.0	Male	34.8
HF	77.9 ± 11.1	Female	36.8
Denoising dataset			
Description			Length(s)
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 tetratone			15
Continuous murmur			15
Overlapping galloping sound			9
Pendulum sound			6
Yaseen dataset			
Description			Number
Aortic Stenosis			200
Mitral Regurgitation			200
Mitral Stenosis			200
Mitral Valve Prolapse			200
Normal			200

2.2. 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.

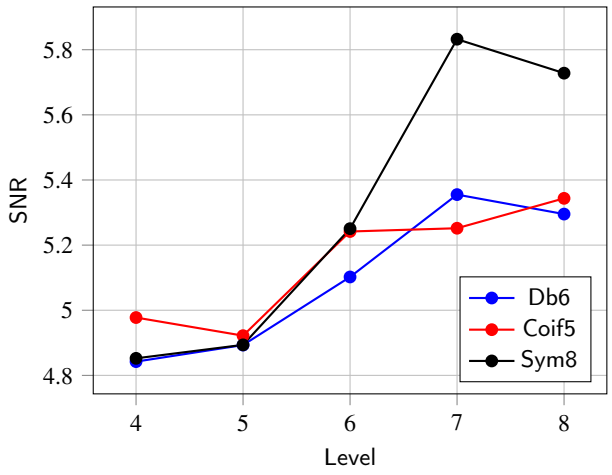
Secondly, we further investigate the average SNR across different shrinkage functions and threshold values while maintaining the sym8 base at the 7-layer decomposition level. Our findings indicate that the use of the 20% modulo maximum with the f_{self} threshold function emerged as the optimal denoising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

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

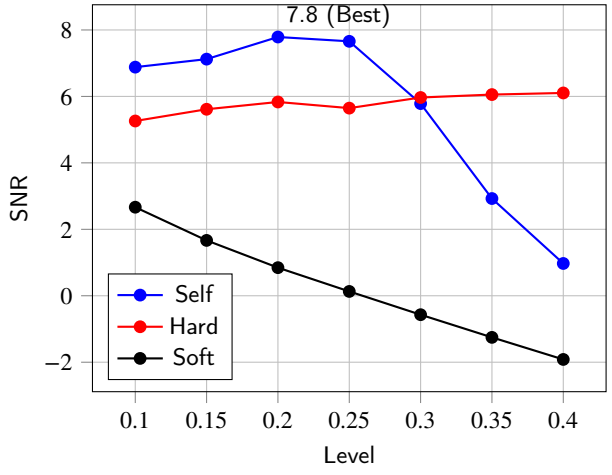
2.3. Multi-region fusion auscultation

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

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



(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}.

$$Se = \frac{TP}{TP + FN} \tag{10}$$

$$Sp = \frac{TN}{TN + FP} \tag{11}$$

$$F1 - Score = 2 \times \frac{Se \times Sp}{Se + Sp} \tag{12}$$

For multi-region fusion auscultation, the average accuracies are 99.35%, 98.71%, and 99.14%, respectively. In terms of average sensitivity, these models achieve 99.08%, 99.09%, and 100.00%, respectively. Furthermore, the average specificity for these models is measured at 99.75%, 98.10%, and 97.95%, respectively. The average F1-Scores are 99.42%, 98.59%, and 98.97%, respectively.

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 exhibits 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%.

2.4. Mitral valve auscultation

Fig.4 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 worth highlighting that both ResNet-18 and MobileNetV1-28 demonstrate similar performance trends, with DenseHF-Net demonstrating the fastest convergence among them.

Tab.3 lists the 10-fold results of HF diagnosis. DenseHF-Net, ResNet-18, and MobileNetV1-28 basically converge around 50 epochs.

For mitral valve auscultation, the average accuracies achieved by the individual models are 93.63%, 91.33%, and 90.73%, respectively. In terms of average sensitivity, these models reach 94.28%, 92.32%, and 90.68%, respectively. Furthermore, the average specificity for these models is measured at 92.95%, 90.20%, and 90.80%, respectively. The average F1-Scores for these models are 93.57%, 91.17%, and 90.69%, respectively.

When combined with DenseHF-Net and the ensemble method, the overall performance improves significantly. The resulting average accuracy, sensitivity, specificity, and F1-score are enhanced to 94.41%, 96.11%, 92.00%, and 93.95%, respectively.

Fig. 3 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 Figures 3a and 3b summarize the performance outcomes, including True Positives, True Negatives, False Positives, and False Negatives.

In the first strategy (Fig. 3a), 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. 3b) demonstrated even higher accuracy, achieving a sensitivity of 98.97% and a specificity of 99.44%, leading to an F1 score of 99.10%.

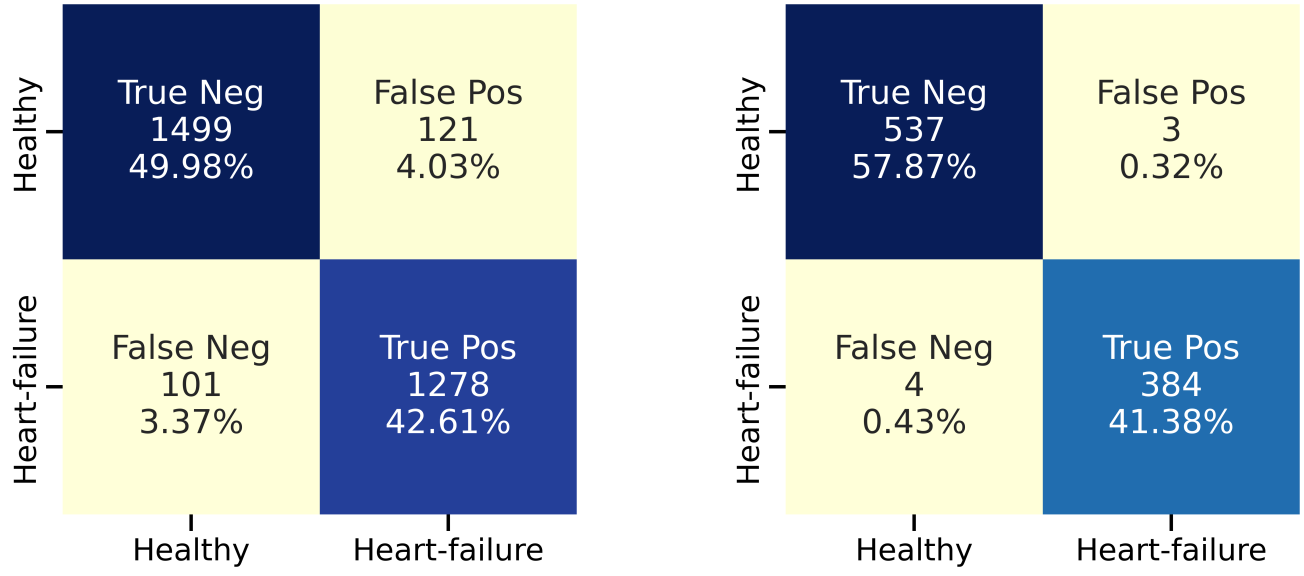
The comparison highlights that while both strategies are effective in detecting mitral valve abnormalities, the second strategy exhibits superior performance across all metrics, particularly in minimizing false negatives and false positives. This suggests that the second strategy may be more reliable for clinical use, where accurate identification of mitral valve conditions is critical for patient outcomes.

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	$Average \pm sd$	MobileNetV1-28	$Average \pm sd$
$Acc(\%)$	99.25 ± 0.71	$Acc(\%)$	98.71 ± 1.26	$Acc(\%)$	99.14 ± 0.65
$Se(\%)$	98.97 ± 1.22	$Se(\%)$	99.09 ± 0.92	$Se(\%)$	100.00 ± 0.00
$Sp(\%)$	99.44 ± 0.75	$Sp(\%)$	98.10 ± 2.68	$Sp(\%)$	97.95 ± 1.54
$F1-Score(\%)$	99.10 ± 0.64	$F1-Score(\%)$	98.59 ± 1.49	$F1-Score(\%)$	98.97 ± 0.78
Mitral valve auscultation results					
DenseHF-Net(ours)	$Average \pm sd$	ResNet-18	$Average \pm sd$	MobileNetV1-28	$Average \pm sd$
$Acc(\%)$	92.60 ± 1.11	$Acc(\%)$	91.33 ± 2.48	$Acc(\%)$	90.73 ± 1.80
$Se(\%)$	92.68 ± 0.88	$Se(\%)$	92.32 ± 2.59	$Se(\%)$	90.68 ± 2.49
$Sp(\%)$	92.53 ± 2.08	$Sp(\%)$	90.20 ± 4.49	$Sp(\%)$	90.80 ± 2.97
$F1-Score(\%)$	92.01 ± 2.15	$F1-Score(\%)$	91.17 ± 2.58	$F1-Score(\%)$	90.69 ± 1.75

Yaseen dataset results

	DenseHF-Net(ours)				ResNet-18				MobileNetV1-28			
	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

(b) Mitral valve auscultation

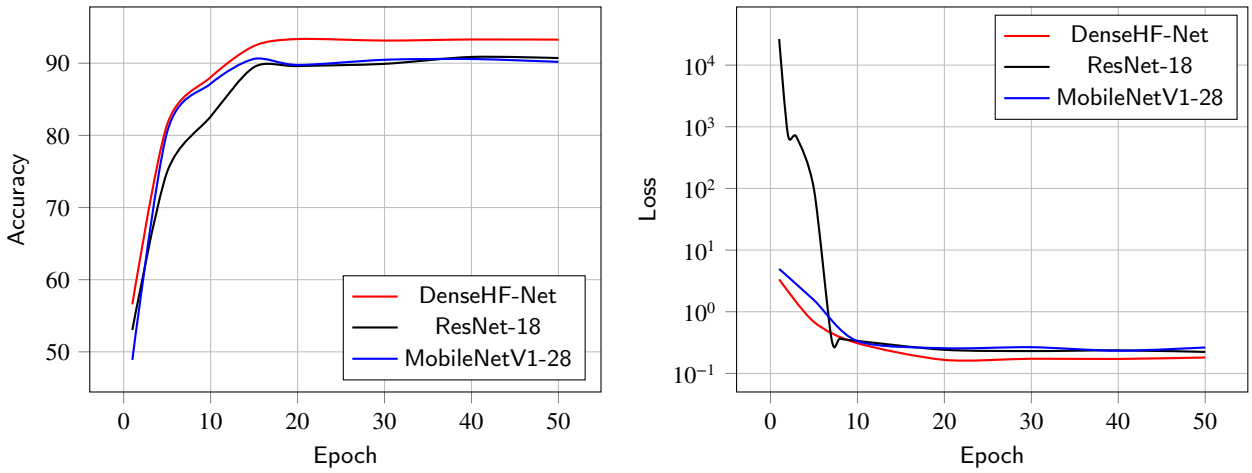
Figure 3: Comparison of two auscultation strategies

2.5. 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.



(a) Mitral valve auscultation testing accuracy.

(b) Mitral valve auscultation testing loss value.

Figure 4: Average 10-fold CV history. (**a**) Average testing accuracy of three models. (**b**) Average testing loss of three models (without augmentation).

Table 4
Ablation Study Results

Ablation Condition	Params (M)	FLOPs (M)	MACs (M)	Accuracy (%)
DenseHF-Net	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

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.

References

[1] G.-Y. Son, S. Kwon, Classification of heart sound signal using multiple features, Applied Sciences 8 (2018) 2344.

[2] 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).

[3] 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.

[4] X. Cheng, Z. Zhang, Denoising method of heart sound signals based on self-construct heart sound wavelet, Aip Advances 4 (2014) 087108.

[5] 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.

[6] 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.

[7] 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.

[8] 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).

[9] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, K. Sricharan, 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.

[10] 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).

[11] 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.

[12] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, A. Gumaei, Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings, IEEE Access 9 (2021) 36955–36967.

[13] 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.

[14] 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.