Heliyon 10 (2024) e35631

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Deep learning fusion framework for automated coronary artery disease detection using raw heart sound signals

YunFei Dai [a,2], PengFei Liu [b,2], WenQing Hou [c,2], Kaisaierjiang Kadier [b], ZhengYang Mu [a], Zang Lu [a], PeiPei Chen [a], Xiang Ma [b,**], JianGuo Dai [a,*,1]

[a] *College of Information Science and Technology, Shihezi University, Shihezi, Xinjiang, 832000, China*
[b] *Department of Cardiology, First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang, 830000, China*
[c] *School of Information Network Security, Xinjiang University of Political Science and Law, Tumushuke, Xinjiang, 843900, China*

ARTICLE INFO

Keywords:
Deep learning
Heart sound signal
Feature fusion
Coronary artery disease
Multidomain features

ABSTRACT

One of the most common cardiovascular diseases is coronary artery disease (CAD). Thus, it is crucial for early CAD diagnosis to control disease progression. Computer-aided CAD detection often converts heart sounds into graphics for analysis. However, this method relies heavily on the subjective experience of experts. Therefore, in this study, we proposed a method for CAD detection using raw heart sound signals by constructing a fusion framework with two CAD detection models: a multidomain feature model and a medical multidomain feature fusion model. We collected heart sound signal datasets from 400 participants, extracting 206 multidomain features and 126 medical multidomain features. The designed framework fused the same one-dimensional deep learning features with different multidomain features for CAD detection. The experimental results showed that the multidomain feature model and the medical multidomain feature fusion model achieved areas under the curve (AUC) of 94.7 % and 92.7 %, respectively, demonstrating the effectiveness of the fusion framework in integrating one-dimensional and cross-domain heart sound features through deep learning algorithms, providing an effective solution for noninvasive CAD detection.

1. Introduction

According to the latest data from the World Health Organization [1], the prevalence and mortality of cardiovascular diseases are continuously rising, making them one of the leading causes of death globally. Coronary Artery Disease (CAD) is the most common among these, with 330 million people currently affected by cardiovascular diseases in China [2], including 11.39 million with CAD. The gold standard for diagnosing CAD is coronary angiography, which is limited by its high equipment costs, expensive medical fees, and invasive nature for patients [3]. Other traditional detection methods also have certain limitations, such as lengthy and costly cardiovascular magnetic resonance imaging [4]; coronary computed tomography angiography (CCTA) has a lower sensitivity of about 78 % when detecting CAD [5]; ultrasound examinations, although non-invasive, cannot directly diagnose [6], and these methods are

* Corresponding author.
** Corresponding author.
E-mail addresses: maxiangxj@yeah.net (X. Ma), djg_inf@shzu.edu.cn (J. Dai).
[1] Lead contact.
[2] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e35631
Received 12 May 2024; Received in revised form 21 June 2024; Accepted 1 August 2024
Available online 3 August 2024

also limited in terms of convenience and detection efficiency. Heart sound analysis, another diagnostic tool, can provide early indications of potential cardiac abnormalities, offering advantages such as non-invasiveness, low cost, and simplicity. However, auscultation depends on the subjective experience of cardiologists, and diagnostic results can vary significantly among experts [7]. To achieve rapid screening and diagnosis of coronary artery disease (CAD), we believe that constructing a CAD detection model based on heart sounds can reduce the reliance on subjective judgments of the experts effectively, improve accuracy and consistency of the diagnosis, make early CAD detection and treatment possible, and improve the prognosis and quality of life of the patient significantly.

Therefore, in this study, we introduced a fusion framework that fuses one-dimensional features separately with multidomain features (MDF) and medical multidomain features (MMDF) derived from heart sound signals. The designed framework resulted in the construction of MDF and MMDF fusion models, enhancing the sensitivity, specificity, and interpretability of the model for CAD diagnosis. Initially, preprocessed heart sound signals are fed into two feature fusion models to extract two types of heart sound signal features: firstly, one-dimensional deep learning features are jointly extracted from heart sound signals using the framework's built-in Convolutional Neural Networks (CNN); Secondly, the MDF and MMDF are extracted from the heart sound signals respectively. Subsequently, deep learning features are separately fused with MDF and MMDF, and then the resulting fused features are fed into the Multilayer Perceptron (MLP) of each model. Features are then mapped to the same dimensionality for fusion and classification. The following are the principal contributions of this study:

We proposed a fusion framework combining the MDF fusion model with deep learning advantages in extracting subtle information with MDF's capability to provide global information. Simultaneously, the MMDF fusion model further enhances interpretability. This framework augments the insight of the model into subtle changes in heart sounds caused by CAD, offering a more precise and comprehensive approach to heart sound signal analysis and CAD detection.

We proposed an innovative one-dimensional convolutional neural network (CNN) structure using one-dimensional heart sound signals as input and integrating an inverse residual structure and a channel attention inversion module (CAIM). This structure directly extracts disease-related information from heart sound signals, avoiding the need for empirical judgments and feature loss associated with converting heart sounds into two-dimensional images.

We tested the heart sound signal data of 400 subjects collected from hospitals, achieving areas under the curve (AUC), sensitivities, and specificities of 94.7 %, 90.7 %, and 82.4 % with the MDF fusion model, respectively, and 92.7 %, 88.0 %, and 80.8 % with the MMDF fusion model, respectively, surpassing the existing research. In the MDF fusion model, a high AUC indicates high accuracy in distinguishing CAD from nonCAD cases, whereas, in the MMDF fusion model, a high AUC provides better interpretability at a lower cost. Overall, this method is an effective, noninvasive approach for CAD detection, which is particularly valuable and promising for areas with limited medical resources and patients unsuitable for invasive testing.

2. Related works

Current research on heart sound signal classification largely relies on public datasets, such as the PhysioNet/CinC Challenge [8–11] and Pascal [8,12,13], with two mainstream approaches: manual feature extraction for machine learning classification and automatic feature extraction and classification via deep learning.

Manual extraction of heart sound features typically involves multiple domains such as time domain [14], frequency domain [15], time–frequency domain [16], wavelet [17], and entropy [18], followed by input into machine learning classifiers such as support vector machines [14,15,19,20], naive Bayes [14,19], k-nearest neighbors [19,20] and random forests [8,19]. For instance, Abduh et al. [16] achieved 92 % mean accuracy by extracting time–frequency domain features and using a support vector machine classifier. Soto-Murillo et al. [19] obtained 80.49 % accuracy by extracting time- and frequency-domain features and applying logistic regression. D. Levin et al. [20] obtained 91 % accuracy by extracting time, frequency, and statistical domain features. Therefore, these studies demonstrate the dependence of traditional manual feature-extraction methods on combining different domain features to reveal disease-related characteristics. However, manual feature extraction is time-consuming, costly, and largely depends on experience and intuition of the researchers, leading to result variability among researchers and affecting consistency and replicability. Furthermore, the performance of CAD detection models will be limited because manually selected features may not capture all critical information in heart sounds. Despite the decent detection performance of traditional manual methods, their limitations have led researchers to seek more automated and efficient feature-extraction approaches.

Deep learning has been successfully applied in multiple fields [21–23]. It has been utilized in recognizing heart sound signals, achieving significant research outcomes [10–12,24–28]. The prevailing approach involves converting one-dimensional heart sound signals into two-dimensional feature maps as input into deep learning models for CAD detection. For example, Zhou et al. [29] achieved 94.1 % sensitivities and 88.7 % specificities by converting one-dimensional heart sound signals into two-dimensional Mel-frequency cepstral coefficient (MFCC) feature maps for classification. Similarly, Chen et al. [26] reached 71.4 % sensitivities and 95.1 % specificities by transforming heart sound signals into two-dimensional Mel spectrum and log-Mel spectrum feature maps. Xiang et al. [25] converted the signals into various two-dimensional forms, including envelope, waveform, log-Mel spectrogram, and log-power spectrogram, which were used as inputs for different models, achieving accuracies of 84.9 % (envelope), 86.6 % (waveform), 93.4 % (log-Mel spectrogram), and 93.8 % (log-power spectrogram) in the InceptionV3 model. The above studies indicate the benefits of high-quality time–frequency diagrams for CAD detection. However, different two-dimensional time–frequency diagrams may vary when capturing the features of raw heart sound signals. Additionally, converting two-dimensional time–frequency images is often based on empirical judgments, which may lead to critical information loss and affect the model's performance. One-dimensional heart sound signals retain more characteristic information from the original heart sounds compared to the time–domain diagram in the two-dimensional diagram. Using one-dimensional heart sound signals as input, combined with different domain features, can effectively prevent information loss during the conversion of two-dimensional time–frequency diagrams, allowing for a more comprehensive capture of features closely related to CAD, improving CAD detection's accuracy and efficiency.

3. Materials and methods

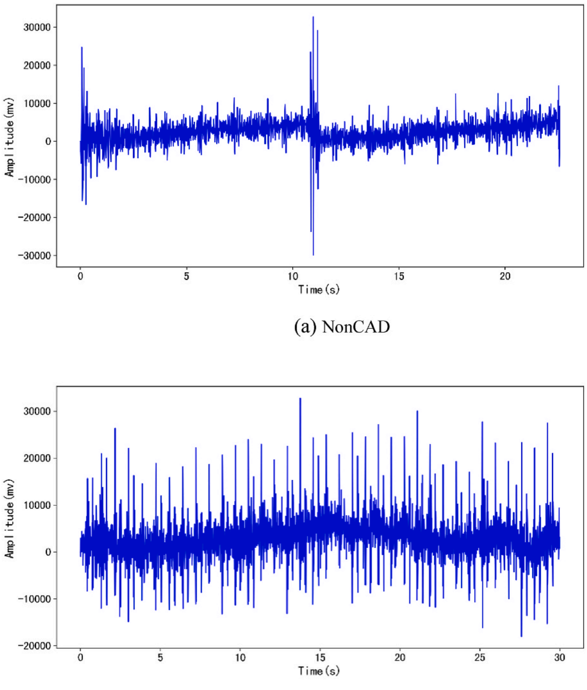
3.1. Data acquisition

This study was performed in accordance with the principles of the Declaration of Helsinki and its amendments and was approved by the ethics committee of Xinjiang Medical University. All participants were patients at the Chest Pain Center of the First Affiliated Hospital of Xinjiang Medical University. Signed informed consent forms were obtained from all participants before participation. The inclusion criteria were patients who underwent coronary angiography or coronary CTA between april 1st and november 1st, 2021, and met the 2013 ESC guidelines for diagnosis and treatment of stable coronary artery disease: aged 18 and above with stable angina symptoms. In this study, participants were required to rest supine for over 10 min to return to normal heart rate; thereafter, heart sounds were recorded at 4000Hz sampling rate at nine different precordial locations using a 3M littmann 3200 electronic stethoscope, for 30-s each. A total of 400 participants were enrolled, including 132 normal individuals and 268 CAD patients, producing 3600 30-s heart sound recordings. CAD was defined as at least one major coronary artery branch (left anterior descending, left circumflex, or right coronary artery) having a narrowing of 50 % or more, with the rest considered normal. Fig. 1 shows a comparative waveforms of heart sounds between CAD patients and normal participants. Compared to normal individuals, CAD patients exhibited significant differences in waveform amplitude, and their heart sounds contained more noise. These noises, represented as atypical waveforms in the figure, reflect cardiac function abnormalities, directly illustrating the impact of CAD on cardiac function.

The locations for the 9-channel heart sound audio collection are as follows: 1. Intersection of the sternal midline with the third intercostal space; 2. Intersection of the sternal midline with the fourth intercostal space; 3. Intersection of the sternal midline with the fifth intercostal space; 4. Intersection of the left sternal border with the third intercostal space; 5. Intersection of the left sternal border with the fourth intercostal space; 6. Intersection of the left sternal border with the fifth intercostal space; 7. 3 cm medially from the midclavicular line at the third intercostal space; 8. 2 cm medially from the midline at the fourth intercostal space; 9. 1.5 cm medially from the midline at the fifth intercostal space.

3.2. Signal processing

We segmented 30-s heart sound signals into several shorter segments for data augmentation to enhance the accuracy of automated heart sound analysis. Given that a cardiac cycle typically lasts about 0.8–1.0s, and having at least three cardiac cycles helps in extracting MDF and MMDF, a duration of 5 s was selected for each segment of the heart sound signal. This approach generated a total of



(a) NonCAD

(b) CAD

Fig. 1. Time–domain waveforms of normal heart sounds and abnormal heart sounds.

21,600 5-s heart sound segments, providing a rich and effective dataset for the study.

Preprocessing of heart sound signals is a critical step in their analysis, aimed at noise reduction and signal quality enhancement to provide a clearer, more accurate data foundation for subsequent analysis. In this study, the Daubechies 8 wavelet function of the wavelet transform was used to perform the noise reduction of heart sound signals, and the number of decomposition layers was set to 10 in order to weaken or eliminate the noise components more effectively at each level and retain the important information in the heart sound signals. Fig. 2 shows the before and after heart sound signal processing.

3.3. *Cardiac cycle segmentation*

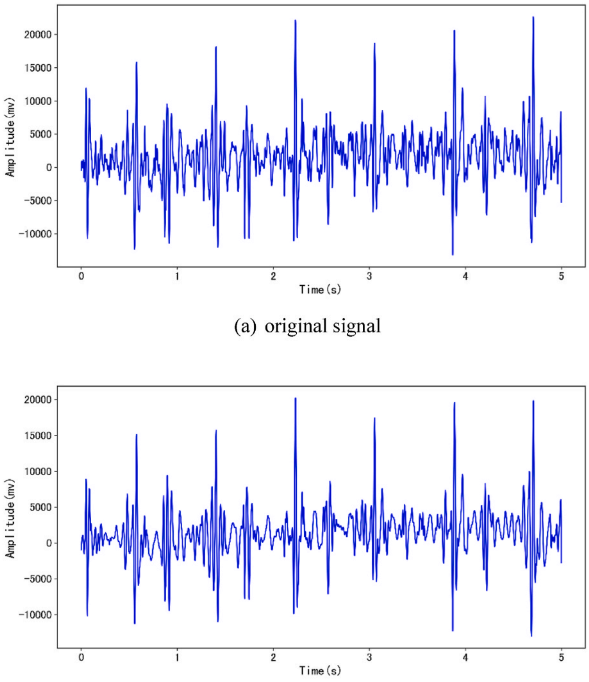
Cardiac cycle segmentation is a crucial step in the clinical diagnosis of cardiac diseases [30], capable of accurately identifying and isolating the first heart sound (S1) and the second heart sound (S2) within the cardiac cycle. S1 typically arises from vibrations caused by the closure of the mitral and tricuspid valves, marking the start of the cardiac systole. S2 is caused by vibrations during the closure of the aortic and pulmonary valves, indicating the beginning of the cardiac diastole. The interval between these two sounds is the systolic phase, and the interval from S2 to the next cycle's S1 is the diastolic phase. These periodic cardiac events are key to understanding cardiac functional status, but their temporal and characteristic variations increase the complexity of heart sound signal analysis. We employed a logistic regression-based hidden semi-Markov model for segmenting heart sound signals within cardiac cycles. Fig. 3 illustrates the segmentation of the cardiac cycle and its four states.

3.4. *MDF and MMDF extraction*

We extracted 12 time–domain features, 8 cardiac cycle spectral features, 6 entropy features, 10 energy ratio features, and 8 statistical features within the segmented cardiac cycles. In addition, we incorporated 8 spectral features and 22 wavelet features from a 5-s heart sound signal into the MDF of the study. We proposed a medically informed MMDF based on coronary physiology [31] encompassing 6 time–domain features, 4 cardiac cycle spectral features, 6 entropy features, 6 energy ratio features, 4 statistical features, and 22 wavelet features from a 5-s heart sound signal.

3.4.1. *Time-domain features (12 × 4Features , 6 × 4Features)*

Time-domain features of normal and abnormal heart sounds may contain characteristics beneficial for the detection of CAD. The MDF calculates the durations of four states within the cardiac cycle: S1, systole, S2, and diastole, as well as the ratios of time and amplitude within the cardiac cycle [32]. Considering factors such as the maximum blood flow velocity in the left coronary artery during diastole and the roughly equal blood flow velocities in the right coronary artery during both systole and diastole, the MMDF computes specific time-domain features, as detailed in Table 1.



(a) original signal

(b) processed signal

Fig. 2. Before and after heart sound signal processing.

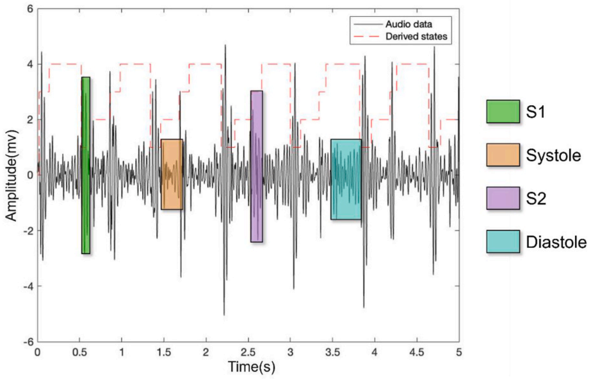


Fig. 3. Periodic segmentation of heart sound signals.

3.4.2. *Cardiac cycle spectral features (8 × 4Features , 4 × 4Features)*

The MDF analyzes spectral features of four states in the cardiac cycle, setting 50Hz as the low-frequency threshold and 200Hz as the high-frequency threshold. Considering that the dominant and central frequencies of heart sound signals typically range from 20 to 40Hz after a fast fourier transform, the MMDF adjusts the low-frequency threshold to 20–150Hz and the high-frequency threshold to 150–1000Hz to calculate features during systole and diastole, as detailed in Table 2.

3.4.3. *Entropy features (6 × 4Features , 6 × 4Features)*

The non-linear and non-stationary characteristics of heart sound signals make traditional linear analysis methods often inadequate for accurately capturing their dynamic features. Entropy features, by measuring the system's uncertainty and complexity, offer deeper insights into the intrinsic dynamics of heart sound signals. Both the MDF and the MMDF analyze three types of entropy features during systole and diastole, as detailed in Table 3.

1. Sample entropy [33] is calculated as follows:

$$SampEn(m,r,r) = -\ln \frac{\sum_{i=1}^{N-m} B_i^{(m+1)}(r)}{\sum_{i=1}^{N-m} B_i^{(m)}(r)} \tag{1}$$

In the equation: N is the length of the heart sound signal, m is the embedding dimension, r is the threshold parameter, and $B_i^{(m)}(r)$ is the probability that any two fixed time Windows match each other.

2. The membership function for the fuzzy similarity of fuzzy entropy between X_i^m and X_j^m is usually chosen as the exponential decay function. The fuzzy entropy [34] is calculated as follows:

$$S_{i,j}^m = e^{\left(-d_{i,j}^2 / r\right)} \tag{2}$$

In the equation: $-d_{i,j}^2$ is the distance between X_i^m and X_j^m, the pattern length of SampEn and FuzzyEn is set to 2, and the matching

Table 1
All time-domain features were extracted from each cardiac cycle.

Feature abbreviation	Description	MMDF
DurCc	Sum of duration of the 4 states of the cardiac cycle	✓
DurS1	The interval from S1 to the next adjacent systolic period	
DurSys	The interval from systolic period to the next adjacent S2	✓
DurS2	The interval from S2 to the next adjacent diastolic period	
DurDia	The interval from diastolic period to the next adjacent S1	✓
SysToCc	Ratio of systolic period duration to cardiac cycle duration	✓
DiaToCc	Ratio of diastolic period duration to cardiac cycle duration	✓
SysToDia	Ratio of systolic period duration to diastolic period duration	✓
SysToS1_Amp	Ratio of average amplitude of systole period to S1	
DiaToS2_Amp	Ratio of average amplitude of diastolic period to S2	
S1ToCc_Amp	Ratio of average amplitude of S1 to cardiac cycle	
S2ToCc_Amp	Ratio of average amplitude of S2 to cardiac cycle	

The MDF are shown in the table. The third column "√" indicates the MMDF.

Table 2
All spectral features were extracted from each cardiac cycle.

Feature abbreviation	Description	MMDF
HF_S1	The proportion of high frequency components in S1	
LF_S1	The proportion of the low frequency component in S1	
HF_S2	The proportion of high frequency components in S2	
LF_S2	The proportion of the low frequency component in S2	
HF_Sys	The proportion of high frequency components in systole period	✓
LF_Sys	The proportion of the low frequency component in systole period	✓
HF_Dia	The proportion of high frequency components in diastole period	✓
LF_Dia	The proportion of the low frequency component in diastole period	✓

The MDF are shown in the table. The third column "$\surd$" indicates the MMDF.

Table 3
All entropy features were extracted from each cardiac cycle.

Feature abbreviation	Description	MMDF
Samp_Sys	Systolic period sample entropy	✓
Samp_Dia	Diastolic period sample entropy	✓
Fuzzy_Sys	Systolic period fuzzy entropy	✓
Fuzzy_Dia	Diastolic period fuzzy entropy	✓
Dist_Sys	Systolic period distribution entropy	✓
Dist_Dia	Diastolic period distribution entropy	✓

The MDF are shown in the table. The third column "$\surd$" indicates the MMDF.

tolerance r is set to 0.2 times the input time series.

3. Distribution entropy [35] is calculated as follows :

$$DistEn(m) = -\frac{1}{\log_2(B)}\sum_{m=1}^{B} p_t \log_2(p_t) \tag{3}$$

In the equation: p_t represents the probability distribution function of random variable t, B represents the bin number of the histogram, and B is set to 2*8 in this study.

3.4.4. Energy ratio features (10 × 4Features , 6 × 4Features)

Energy ratio features are used to assess the energy levels of heart sound signals. The MDF analyzes the energy ratios of different cardiac cycle states to the entire cardiac cycle, as well as the energy ratios between two states. Considering potential diastolic murmurs caused by CAD and possible reductions in the amplitudes of S1 and S2, the MMDF calculates the characteristics of S1, S2, and the diastolic period, as detailed in Table 4. The method for calculating energy is as follows:

$$Energy = \sum_{i=1}^{N} X_i^2 \tag{4}$$

In the equation: i represents discrete time points, N represents the length of the signal X_i.

Table 4
All energy ratio features were extracted from each cardiac cycle.

Feature abbreviation	Description	MMDF
En_S1Sys	The ratio of S1 energy to systolic period energy	
En_S1Dia	The ratio of S1 energy to diastole period energy	✓
En_S2Sys	The ratio of S2 energy to systolic period energy	
En_S2Dia	The ratio of S2 energy to diastole period energy	
En_DiaSys	The ratio of diastole period energy to systolic period energy	✓
En_S1Cc	The ratio of S1 energy to cardiac cycle energy	✓
En_SysCc	The ratio of systolic period energy to cardiac cycle energy	
En_S2Cc	The ratio of S2 energy to cardiac cycle energy	✓
En_DiaCc	The ratio of diastole period energy to cardiac cycle energy	✓
En_S1S2Cc	The ratio of the sum of S1 and S2 energies to cardiac cycle energy	✓

The MDF are shown in the table. The third column "$\surd$" indicates the MMDF.

3.4.5. Statistical features (8 × 4Features , 4 × 4Features)

In the analysis of heart sound signals, skewness and kurtosis are used to assess the asymmetry and peak distribution characteristics of heart sound signals at different stages of the cardiac cycle [36]. The MDF calculates the skewness and kurtosis for the four states of each cardiac cycle. The MMDF calculates these statistical features for S1 and S2. The calculations of skewness and kurtosis are as follows:

$$skewness = \frac{\sum_{i=1}^{N}(x_i - m)^3}{N \times sd^3} \tag{5}$$

$$kurtosis = \frac{\sum_{i=1}^{N}(x_i - m)^4}{N \times sd4} - 3 \tag{6}$$

In the equation: m and sd represent the mean and standard deviation, respectively, of the time series x_i.

3.4.6. Heart sound spectral features (8 × 1Features , 0 × 0Features)

By analyzing spectral information such as frequency components and energy distribution in heart sound signals, spectral features related to CAD can be extracted, which are beneficial for CAD detection. The MDF sets 50Hz and 200Hz as the low and high frequency thresholds, respectively, to calculate the spectrum of the entire heart sound signal. The MMDF chooses not to compute this feature, as detailed in Table 5. The entropy calculation is as follows:

$$Entropy = -\sum_{i=1}^{L}\frac{e_i}{\sum_{j=1}^{L}e_j} \times \log\frac{e_i}{\sum_{j=1}^{L}e_j} \tag{7}$$

In the equation: L represents the sequence length.

3.4.7. Wavelet features (22 × 1Features , 22 × 1Features)

Wavelet transform [37] is an effective time-frequency analysis tool. Both MDF and MMDF use daubechies 6 wavelets for a three-level decomposition at a 2000Hz sampling rate of heart sound signals, calculating approximation and detail coefficients to derive nine wavelet features, including energy across various frequency bands and total energy. For an in-depth analysis of high-frequency and low-frequency parts within a 5-s heart sound signal, an eight-vector reconstruction is carried out through a third-level wavelet packet decomposition, extracting thirteen wavelet packet features, including total energy, energy across different frequency bands (formula 1), and energy entropy (formula 7), providing a comprehensive perspective on high-frequency analysis of heart sound signals. Table 6 details the specific wavelet feature information.

3.5. One-dimensional convolutional neural network model

3.5.1. Benchmark model

While small convolution kernel in deep learning models can capture local detail features of heart sound signals, they may not fully comprehend global information when processing heart sound signals. To address this limitation, this study introduces the Inception module from GoogleNet [38], which is equipped with convolution kernel of multiple sizes to effectively capture heart sound features at various scales, thereby enhancing the accuracy of CAD detection. However, increased model depth leads to the problem of vanishing gradients. Thus, referencing the residual design of ResNet [39], this study uses skip connections to allow gradients to flow directly through certain layers, effectively mitigating the problem of deep network gradient vanishing in the CAD detection model while also accelerating the model's convergence process.

Fig. 4 shows the benchmark model, which includes inputs, outputs, inception blocks, and shortcut blocks. The inception blocks internally integrate multiscale convolution and a CAIM. In this study, the heart sound signals with a duration of 5-s were resampled at 2000 Hz, resulting in a total of 10,000 samples. The kernel sizes of traditional Inception modules (1, 3, 5) are mainly used for image processing and are not effective for extracting features from heart sound signals with many samples. Therefore, when designing the

Table 5
All spectral features were extracted from each 5-s heart sound signal.

Feature abbreviation	Description	MMDF
Mean_spec	The average spectrum value of the entire heart sound signal	
Sd_spec	The standard deviation of the spectral values of the entire heart sound signal	
Ske_spec	The skewness of the spectral values of the entire heart sound signal	
Kur_spect	The kurtosis of the spectral values of the entire heart sound signal	
HighF_ratio	The proportion of high-frequency components in the spectral values of the entire heart sound signal	
LowF_ratio	The proportion of low-frequency components in the spectral values of the entire heart sound signal	
HighToLow _ratio	The ratio of high-frequency to low-frequency components in the spectral values of the entire heart sound signal	
Ep_spec	The entropy of the spectral values of the entire heart sound signal	

The MDF are shown in the table. The third column "√" indicates the MMDF.

Table 6
All wavelet features were extracted from each 5-s heart sound signal.

Feature abbreviation	Description	MMDF
Wdec_ratio1	Wavelet approximation coefficient energy in the range 0–62.5 Hz for the whole heart sound band	✓
Wdec_ratio2	Wavelet detail coefficient energy in the range 62.5–125 Hz for the whole heart sound band	✓
Wdec_ratio3	Wavelet detail coefficient energy in the range 125–250 Hz for the whole heart sound band	✓
Wdec_ratio4	Wavelet detail coefficient energy in the range 250–500 Hz for the whole heart sound band	✓
Wdec_ratio5	Wavelet detail coefficient energy in the 500–1000 Hz range for the whole heart sound band	✓
Wdec_total	Total wavelet coefficient energy in the 0–1000 Hz range for the whole heart sound band	✓
Wdec_l_ratio	Total energy of wavelet coefficients below 125 Hz for the whole heart sound band	✓
Wdec_m_ratio	Total energy of wavelet coefficients in the range 125–500 Hz for the whole heart sound band	✓
Wdec_h_ratio	Total energy of wavelet packet decomposition features above 500 Hz for the whole heart sound band	✓
Wpac_ratio1	Reconstruction coefficient energy in the 0–125 Hz range across the heart sound	✓
Wpac_ratio2	Reconstruction coefficient energy in the 125–250 Hz range across the heart sound	✓
Wpac_ratio3	Reconstruction coefficient energy in the 250–375 Hz range across the heart sound	✓
Wpac_ratio4	Reconstruction coefficient energy in the 375–500 Hz range across the heart sound	✓
Wpac_ratio5	Reconstruction coefficient energy in the 500–625 Hz range across the heart sound	✓
Wpac_ratio6	Reconstruction coefficient energy in the 625–750 Hz range across the heart sound	✓
Wpac_ratio7	Reconstruction coefficient energy in the 750–875 Hz range across the heart sound	✓
Wpac_ratio8	Reconstruction coefficient energy in the 875–1000 Hz range across the heart sound	✓
Wpec_total	Total energy of reconstructed coefficients in the 0–1000 Hz range of the entire heart sound	✓
Wpac_l_ratio	Total energy of reconstruction coefficients below 125 Hz for the whole heart sound segment	✓
Wpac_m_ratio	Total energy of reconstruction coefficients in the 125–500 Hz range across the heart sound segment	✓
Wpac_h_ratio	Total energy of reconstruction coefficients above 500 Hz for the whole heart sound band	✓
Wpac_entropy Energy	Energy entropy of reconstructed coefficients in eight frequency bands of the whole heart sound band	✓

The MDF are shown in the table. The third column "√" indicates the MMDF.

Inception module for heart sound signal analysis, we set the kernel sizes to k, k/2, and k/4 based on the sampling frequency and feature extraction requirements of heart sound signals. In this study, we set k to 40, which means the kernel sizes are 40, 20, and 10, respectively. This multi-scale kernel setting captures features at various scales in the signal, extracting rich feature information and enhancing the model's ability to analyze heart sound signals. CAIM is specifically designed to enhance the processing effect after each convolution and pooling, aiming to ensure the model fully utilizes heart sound features extracted from various levels, thus improving the efficiency of extracting features beneficial for CAD detection. The input to the benchmark model is a heart sound signal, which is processed in the Inception block numbered 1. Each Inception block is numbered in ascending order, and the heart sound signal passes through these blocks sequentially, with each block processing the signal and feeding it into the next. There is a specific rule for the output of Inception blocks: if the block's number is not a multiple of three, the output remains unchanged; if it is a multiple of three, the output is determined by the Shortcut block. The Shortcut block's function is to sum the current Inception block's output with a residual (initially set as the input's original heart sound features) and update the residual to the Shortcut's output result. Finally, Dropout is placed after multi-scale convolution to improve model generalization and reduce the risk of overfitting.

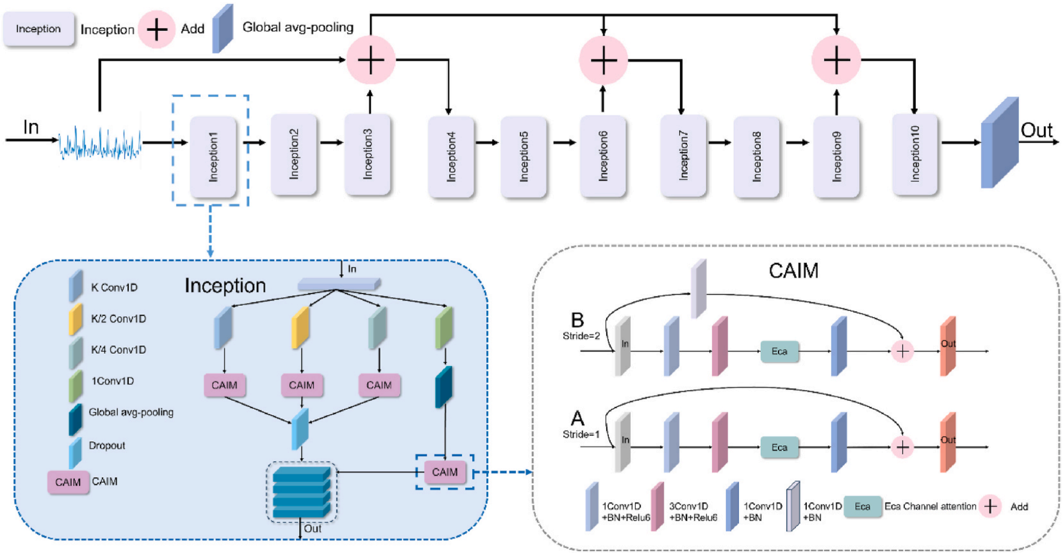


Fig. 4. 1D CNN model.

3.5.2. Channel attention inversion module

The inverted residual module proposed by MobileNetV2 [40] first expands and then compresses the feature channels, which effectively enhances the network's ability to extract heartbeat features. However, the inverted residual module cannot adequately distinguish the importance of different channels, resulting in less accurate heartbeat feature extraction.

Wang et al. proposed the Efficient Channel Attention (ECA) module [41], which is able to highlight important heart sound features more accurately and can ignore unimportant information.

To fully leverage the excellent capability of the inverted residual module in deep feature mining and further enhance the performance of the CAD detection model, we introduced CAIM (as shown in Fig. 4) by integrating the ECA channel attention mechanism. CAIM combines two advantages: first, the expand-compress strategy of the inverted residual module effectively reduces the number of parameters in the CAD detection model while retaining excellent heart sound feature capture capability. Second, by incorporating the ECA module, the model precisely allocates attention weights at minimal cost, thereby making key heart sound features more prominent and improving the accuracy of the CAD detection task.

First, according to equation (8), independent weight multiplication operations are performed on each input channel. The input tensor $X_{l,c} \in R^{C\times L}$ (where L represents the length of the one-dimensional features, and C represents the number of channels) is multiplied with the one-dimensional convolution kernel $W_{c,c'}$ for each input channel c. The resulting products are summed to obtain the output tensor $Y_{l,c'} \in R^{2C\times L}$. In this summation, the convolution kernel has a size of 1, and the number of channels is increased from C to 2 C, achieving an increase in dimensionality. Subsequently, Batch Normalization (BN) (equation (9)) and the ReLU6 activation function (equation (10)) are applied.

$$Y_{l,c'} = \sum\nolimits_{c=1}^{C} X_{l,c} \cdot \quad W_{c,c'}, c' = 1, 2\ldots, 2C \tag{8}$$

In the equation: $X_{l,c}$ is the input feature, $W_{c,c'}$ is the weight of the convolution kernel. l is the index of the feature vector, ranging from 1 to L, while c and c' are the indices of the input and output channels, respectively, with subsequent convolution operations denoted by ©.

$$BN = \frac{Y_{lc'} - \mu_{c'}}{\sqrt{\sigma_{c'}^2 + \varepsilon}} \quad \cdot \quad \gamma_{c'} + \beta_{c'} \tag{9}$$

In the equation: $Y_{lc'}$ is the input tensor, $\mu_{c'}$ is the mean of the c' channel, $\sigma_{c'}^2$ is the variance of the c' channel, ε is a small constant to prevent division by zero, $\gamma_{c'}$ and $\beta_{c'}$ are the learnable parameters for scaling and shifting, with subsequent operations denoted by ® representing the ReLU6 function.

$$y = \mathrm{Re}LU6(x) = \min(\max(x, 0), 6) \tag{10}$$

In the equation: ∅ subsequently represents the BN operation.

Subsequently, according to equation (11), the tensor is slid across each channel, taking into account the current position and the surrounding context. The tensor $Y_{l,c'}$ from the previous step is subjected to feature extraction using a one-dimensional convolution of size 3, resulting in the output tensor $Y_1 \in R^{2C\times L}$. This step is followed by batch normalization and ReLU6 activation:

$$Y_1 = \sum\nolimits_{i=-1}^{1} \sum\nolimits_{c=1}^{2C} Y_{l+i,c} \cdot \quad W_{i,c,c'} \tag{11}$$

In the equation: $\sum_{i=-1}^{1}$ is the index along the spatial dimension, ranging from -1 to 1, representing a convolution kernel size of 3. $\sum_{c=1}^{2C}$ is the index along the channel dimension of the input feature map, corresponding to all channels at each position. $Y_{l+i,c}$ is the tensor of the input feature map at position $l+i$ and channel c. $W_{i,c,c'}$ is the weight of the convolution kernel from input channel c to output

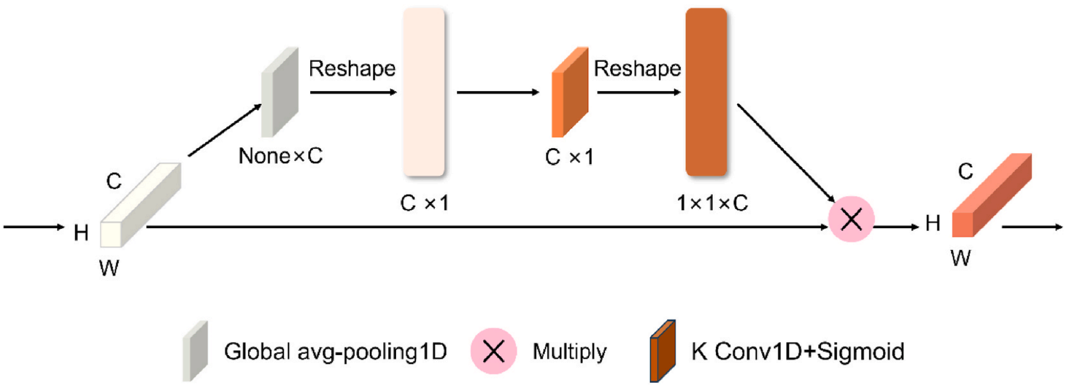


Fig. 5. ECA module.

channel c'. i ranges from -1 to 1, representing the current and adjacent positions. l is the spatial location of the currently processed feature map, c is the index of the input channel, and c' is the index of the output channel.

Subsequently, tensor $Y_1 \in R^{2C\times L}$ enters the ECA module (as shown in Fig. 5). ECA focuses solely on the interaction of the input tensor Y_1 with its k adjacent neighboring channels.

In ECA, the input feature map is first subjected to global average pooling, followed by a one-dimensional convolution with a kernel size of K, and the weights ω for each channel are obtained through a Sigmoid activation function, as shown in equation (12):

$$\omega = \sigma(C1D_k(y)) \tag{12}$$

In the equation: $C1D_k$ represents a one-dimensional convolution with a kernel size of K ($1 \times k$) and σ denotes the sigmoid function.

The size of the convolution kernel k can be adaptively determined by equation (13).

$$k = \psi(C) = \left|\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{13}$$

In the equation: represents the odd number closest to the internal number.

In ECA, the channel attention weights for $Y_1' \in R^{2C\times L}$ are obtained. According to equation (14), the input tensor $Y_1 \in R^{2C\times L}$ is multiplied by the attention weights to produce the output tensor $Y_2 \in R^{2C\times L}$. This process aims to allocate higher weights to more critical features in the current CAD detection process, thus enhancing the model's performance.

$$Y_2 = Y_1 \quad \cdot \quad Y_1' \tag{14}$$

After the ECA module, the tensor $Y_2 \in R^{2C\times L}$ undergoes reduction and batch normalization using a one-dimensional convolution of size 1. According to equation (15), the output tensor is $Y_3 \in R^{C\times L}$.

$$Y_3 = ®\left(\varnothing\left(Y_2 © K_{1,1}\right)\right) \tag{15}$$

In the equation: $K_{1,1}$ represents a one-dimensional convolution with a size of 1.

Finally, residual connections between A and B in CAIM as shown in Fig. 4. When the stride is 1 and the number of input and output feature channels matches, the network directly adds the output of the current ECA module to the network's original input. In cases where the stride is 2 or the feature channels do not match, the network first adjusts the size and number of channels of the input through a one-dimensional convolution and batch normalization to match the output of the ECA module.

In summary, CAIM has the following advantages over ECA:

(1) Enhanced feature expression: ECA primarily recalibrates features through the weights of each channel. In contrast, CAIM incorporates the inverted residual structure, allowing for more effective transmission and enhancement of feature expression in the convolutional layers, enabling the model to capture more subtle features of heart sound signals.

(2) Reduced feature loss: In CAIM, an attention mechanism is introduced between 1D convolutions of size 3 and size 1, which better preserves and enhances features during the convolution process, reducing information loss during transformation. ECA mainly focuses on feature recalibration within a single layer.

3.6. Feature fusion framework

We proposed a fusion framework (as shown in Fig. 6) to separately integrate MDF and MMDF with the deep learning features of

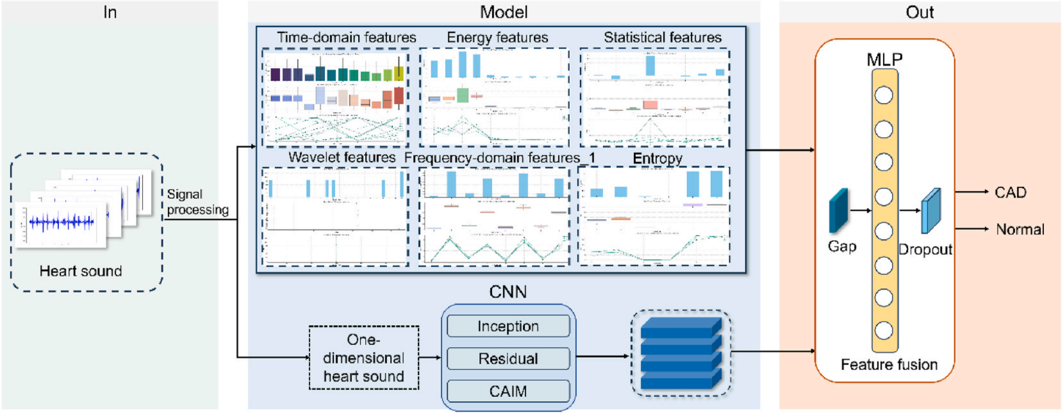


Fig. 6. A coronary artery disease detection model that fuses one-dimensional deep learning features with multi-domain features extracted from heart sound signals.

one-dimensional heart sound signals and construct both an MDF fusion model and an MMDF fusion model. In this framework, the MDF fusion model and MMDF fusion model take preprocessed one-dimensional heart sound signals as inputs and use a one-dimensional CNN structure to extract CAD-related deep features. Each model extracts MDF or MMDF from temporal, frequency, and other key domains to ensure comprehensive heart sound analysis. Finally, the deep learning features were separately fused with MDF and MMDF in the same manner and then inputted into the MLP of the two models for further processing. Subsequently, a specific fusion strategy merges features of the same dimension to form a comprehensive feature representation. We applied the dropout method to the fusion features of both models separately to prevent overfitting and optimize the model's performance. Ultimately, we mapped the features to binary probabilities using the sigmoid activation function, thus realizing CAD detection using the MDF and MMDF fusion models.

3.7. Experimental platform

The parameters used in model training and verification are shown in Table 7.

3.8. Evaluation metrics

We used evaluation metrics to comprehensively evaluate the classification results, such as accuracy, sensitivity, specificity, receiver operating characteristic curve, and AUC value. In the ROC curve, the horizontal coordinate is the false positive rate (FPR), and the vertical coordinate is the true positive rate (TPR), with AUC defined as the area under the ROC curve. For binary classification problems, samples are categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The calculations for these evaluation metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{16}$$

$$Sensitivity = \frac{\mathrm{TP}}{TP + FN} \tag{17}$$

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

$$FPR = \frac{FP}{FP + TN} \tag{19}$$

$$TRP = \frac{TP}{TP + FN} \tag{20}$$

4. Results

4.1. Study population

After removing data from unavailable patients, 396 patients were divided into training, validation, and test sets in a 6:2:2 ratio. We strictly adhered to the principle that each subject's heart sound signals participate in only one dataset, ensuring no cross-contamination of signals between different datasets. To optimize model performance and evaluate its generalization capability, we employed a 5-fold cross-validation strategy. During training, the training set was further divided into 5 subsets. In each iteration, 4 subsets were used for training and 1 subset for validation. This process was repeated 5 times, with each subset used as the validation set once. From these 5 models, we selected the one that performed best on the validation sets and applied it to an independent test set for final evaluation. In the training set, 159 patients (67.09 %) had CAD, of which 124 were male (77.99 %); in the validation set, 53 patients (67.09 %) had CAD, with 39 males (73.59 %); in the test set, 53 patients (66.93 %) had CAD, with 43 males (81.13 %). The baseline characteristics of patients in the training, validation, and test sets are summarized in Table 8.

We found a significant difference in the history of hyperlipidemia ($p < 0.05$) when comparing the baseline characteristics between

Table 7
Experimental platform details.

Environment	Details
GPU	NVIDIA GeForce RTX 3090
CPU	Intel i9-10900X
CPU core	10
Display memory	24 GB
Memory	64 GB
Operating system	Ubuntu 18.04
CUDA	Cudnn-8.0
Python	3.8.0
Deep learning Framework	TensorFlow

the training and validation sets. Additionally, we observed a significant difference in the incidence of stroke between the training and test sets ($p < 0.05$).

4.2. Quantitative analysis

In this study, multiple repetitions were conducted to ensure the reliability and reproducibility of the experimental results. The proposed fusion framework includes an MDF fusion model and an MMDF feature fusion model, which were compared with recently outstanding classification models in the field of heart sounds, including ResNet18, BotNet, ResNet50, CotNet50, InceptionV4, and MobilenetV3, all designed for processing two-dimensional feature maps. To ensure consistency and fairness in experimental conditions, these models were fed with uniformly sized, unprocessed heart sound time-domain waveforms, while our model, a one-dimensional convolution fusion model, used one-dimensional heart sound time-domain signals. Table 9 shows the comparative results. The proposed MDF fusion model achieved an accuracy of 87.86 %, sensitivity of 90.67 %, specificity of 82.38, and AUC of 94.7 %, while the MMDF fusion model had an accuracy of 85.60 %, sensitivity of 88.04 %, specificity of 80.83 %, and AUC of 92.74 %. Both models outperformed the others on all evaluation metrics, demonstrating superior robustness and accuracy.

As shown in Table 9, other models exhibit considerable fluctuation in sensitivity and specificity, particularly with sensitivity being much higher than specificity, indicating a higher sensitivity to CAD. However, this leads to a higher rate of false positives when distinguishing between normal heart sounds and CAD heart sounds. Converting heart sound signals into two-dimensional time-domain images is a common practice, but this process may distort or lose some key features, and the newly generated two-dimensional images do not significantly enhance feature representation, thereby affecting the classification of CAD. Converting heart sound signal into time-frequency diagrams is a complex process that requires empirical judgment. The conditions under which heart sound are collected vary from person to person, such as equipment, external noise, and operating methods. This variability increases the difficulty of accurately converting heart sound signals into high-quality time-frequency diagrams. Improper handling of this conversion can negatively impact subsequent CAD detection.

In summary, the fusion model proposed in this study demonstrates significant robustness and accuracy in classifying heart sound signals, particularly in capturing CAD pathological information from signals with complex temporal structures, thus avoiding information loss during the conversion to two-dimensional image.

4.3. Ablation experiments

In this study, dropout ablation experiments were conducted on a one-dimensional convolution neural network model to evaluate the role of dropout in CAD detection tasks. As shown in Table 10, the experiment compared the original model ("Origin") without dropout and the base model ("Base") with dropout. The results indicate that the "Base" model with dropout significantly outperforms the original model, achieving accuracies, sensitivities, specificities, and AUC of 78.7 %, 77.7 %, 80.7 %, and 88.3 %, respectively. This demonstrates that dropout not only reduces overfitting but also enhances the generalization ability of the CAD detection model.

To validate the effectiveness of the CAIM module, we conducted ablation experiments on the “ECA” and “inverted residual” components of CAIM using the “base” model. Table 11 shows that both “ECA” and “inverted residual” improved the accuracy and sensitivity of the model. The proposed CAIM significantly increased the model’s AUC by combining “ECA” and “inverted residual,” demonstrating that CAIM enhances the classification performance of the model while reducing the likelihood of misclassifying healthy individuals as CAD patients, reflecting the innovation and effectiveness of the proposed module.

Using CAIM, MDF, and MMDF, we conducted ablation experiments on the two CAD detection models of the feature fusion framework to assess their contribution to CAD detection. As shown in Table 12, demonstrated excellent performance in CAD detection, with an accuracy of 78.7 % and an AUC of 88.3 %, proving the effectiveness of one-dimensional heart sound signals as inputs for CAD detection. After integrating the "CAIM" block, the model’s accuracy increased to 83.5 %, and AUC to 90.7 %, indicates that the “CAIM”

Table 8
[42] Overview of patient data in training, validation, and test sets.

Characteristic	Training set (n = 237)	Validation set (n = 79)	P^a-value	Test set (n = 80)	P^b-value
Age (Years)	58.00 ± 10.77	59.00 ± 11.40	0.41	58.00 ± 11.83	0.83
Sex (Male)	159 (67.09 %)	50 (63.29 %)	0.63	55 (68.75 %)	0.89
BMI(kg/m^2)	25.81 (25.00,4.00)	25.49 (25.00,4.50)	0.52	26.30 (26.00,6.00)	0.54
EF (Ejection Fraction)	60.84 (62.68,3.32)	60.36 (67.90,3.39)	0.24	61.65 (62.68,2.01)	0.52
NT-Pro BNP	389.28 (93.00,204.40)	212.96 (89.2152.05)	0.72	189.72 (80.75,136.18)	0.25
Troponin	0.097 (0.012,0.005)	0.106 (0.012,0.005)	0.97	0.067 (0.012,0.000)	0.25
Previous history					
Hypertension	99 (41.77 %)	38 (48.10 %)	0.39	32 (40.00 %)	0.88
Diabetes	59 (22.89 %)	16 (20.25 %)	0.49	15 (18.75 %)	0.33
Hyperlipidemia	3 (1.27 %)	0 (0 %)	<0.05	2 (2.5 %)	0.80
Stroke	4 (1.69 %)	3 (3.80 %)	0.50	6 (7.5 %)	<0.05

Data are presented as mean ± standard deviation, mean (median and interquartile range), and count (n%). Due to the small values, troponin is represented with three decimal places, while other values are shown with two decimal places. BMI, Body Mass Index; EF, Ejection Fraction; NT-Pro BNP, N-terminal pro-brain natriuretic peptide; P^a values are obtained by comparing the training and validation groups. Comparison between the training and test groups yields the P^b value.

Table 9
Conducted classification experiments on numerous models.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
ResNet18 [39]	68.35	63.62	56.41	54.70
BoTNet [43]	70.32	83.33	43.95	60.52
ResNet50 [39]	70.32	82.70	45.22	63.96
CotNet50 [44]	75.37	86.16	53.50	69.83
InceptionV4 [45]	70.61	85.30	40.89	61.00
MobileNetV3 [46]	78.22	90.94	52.46	71.70
Ours (MDF)	87.86	90.67	82.38	94.70
Ours (MMDF)	85.60	88.04	80.83	92.74

Table 10
Ablation experiment results on Dropout.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
Origin	73.221	83.076	53.754	75.367
Base	78.667	77.659	80.706	88.339

Table 11
Ablation experiment results on ECA and inverted residual module.

ECA	inverted residual	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
×	×	78.667	77.659	80.706	88.338
✓	×	80.937	86.131	70.751	87.750
×	✓	82.405	85.628	75.882	88.987
✓	✓	83.453	85.628	79.049	90.727

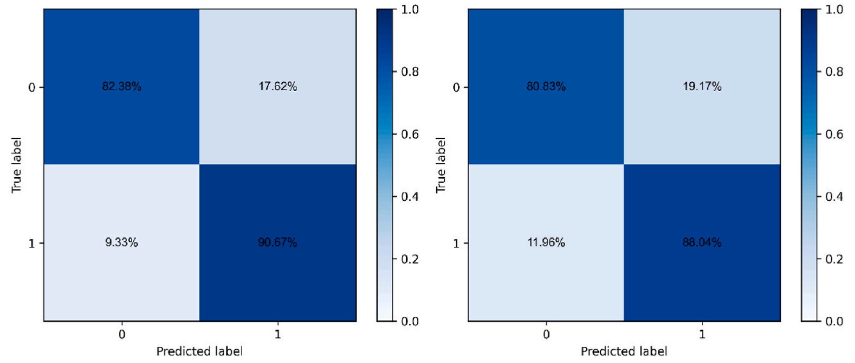
block significantly enhances classification performance. Although specificity slightly decreased, sensitivity significantly improved, reflecting the “CAIM” advantage of the block in enhancing CAD recognition capabilities at a lower cost.

After fusing MDF into the "Base" model, the accuracy increased to 86.5 %, and the AUC significantly rose to 93.2 %. When MMDF were fused, the accuracy and AUC were enhanced to 84.7 % and 89.9 %, respectively. When the "Base" model fused both CAIM blocks and MDF, the accuracy exceeded 87.9 %, indicates that CAIM blocks can effectively extract richer CAD information on top of MDF. Upon integration of CAIM blocks and MMDF into the "Base" model, the accuracy and AUC improved to 85.6 % and 92.7 %, respectively. The results indicate that CAIM blocks can effectively extract richer CAD information than MDF. The results indicated that although the MMDF fusion model performed slightly below the MDF fusion model, MDF significantly enhanced the performance of the model. With only half as many features as MDF, MMDF provided considerably less information to the fusion model, creating this disparity. However, it offered greater medical interpretability. To visually demonstrate the differences between the MDF fusion model and the MMDF model, the confusion matrices of both fusion models were analyzed on the test set. The diagonal values of the confusion matrices represent the percentage of correct classifications for each category. As shown in Fig. 7, using MDF significantly enhanced the model’s ability to distinguish between CAD and normal individuals. In contrast, the detection capability of the MMDF fusion model was somewhat lacking, making it more prone to misclassifying CAD as normal. Nevertheless, due to its structural features, the MMDF strategy provides greater medical interpretability.

In exploring and analyzing the impact of MDF and MMDF on model performance, the "Base" model integrated with CAIM blocks was used to perform ablation studies on MDF and MMDF separately. The results (as in Tables 13 and 14) indicate significant differences in the contribution of features from different domains. In MDF, Time features performed worse than in MMDF, suggesting that amplitude ratio features impacted the performance of the CAD detection model. The cardiac cycle spectral features performed consistently in both MDF and MMDF, indicating potential redundancy in MDF. The Energy features in both MDF and MMDF performed poorly, potentially having a negative impact on CAD classification. In both MDF and MMDF, the Wavelet features exhibited excellent performance, indicating they contain many characteristics beneficial for CAD classification.

Table 12
Ablation experiment results on CAIM, MDF, and MMDF.

CAIM	MDF	MMDF	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
×	×	×	78.667	77.659	80.706	88.338
✓	×	×	83.453	85.628	79.049	90.727
×	✓	×	86.506	89.985	79.719	93.203
×	×	✓	84.745	91.227	72.065	89.866
✓	✓	×	87.861	90.668	82.383	94.704
✓	×	✓	85.599	88.036	80.832	92.741



(a)   (b)

Fig. 7. Indicators of the fusion model across categories (a)Confusion matrix for MDF fusion model; (b) Confusion matrix for MMDF fusion model.

MFCC better aligns with human auditory perceptions of audible sounds. Here, we extracted and analyzed the contribution of MFCC features to CAD detection. Table 16, shows that MFCC features positively impact CAD detection, providing rich spectral information and enabling the model to detect CAD more effectively. Therefore, our fusion framework can efficiently combine useful features for CAD classification. By integrating these features, our model can detect CAD more efficiently, improving classification accuracy.

4.4. *Model interpretability analysis based on SHAP algorithms*

We used the Shapley additive explanations (SHAP) algorithm for interpretability analysis to enhance the interpretability of the model. We computed the Shapley value matrix corresponding to the raw heart sounds using the SHAP gradient explainer and fusion framework model, thus reflecting their contribution to CAD detection. Fig. 8 shows fewer regions with significant waveform changes in normal heart sound signals. The model primarily focuses on the upper envelope, where the waveform changes smoothly and stabilizes rapidly after fluctuations. In CAD heart sound signals, the model pays more attention to regions with significant waveform changes, and these regions typically exhibit a stepped, gradual decline in the subsequent parts. These features help the model identify potential cardiac abnormalities and distinguish normal heart sound signals.

4.5. *Comparison with existing research*

Table 17 showcases prior studies on CAD detection using heart sounds, such as the study by Griffe et al. [47], who achieved an accuracy of 81 % using a support vector machine with heart sound signals from only 31 subjects. Pathak et al. [48] involved heart sound signals from 80 subjects in their study, reaching an accuracy of 84.19 %. Compared to existing single-channel studies, both the MDF and MMDF fusion models in this research achieved superior performance. The MDF fusion model attained an accuracy of 87.9 %, with sensitivity and specificity of 90.7 % and 82.4 %, respectively. The MMDF fusion model recorded an accuracy of 85.6 %, with sensitivity and specificity of 88.0 % and 80.8 %, respectively. This indicates that the fusion models proposed in this study can effectively differentiate between CAD and non-CAD cases.

5. Discussion

5.1. *Fusion framework*

CAD patients experience vascular constriction [18], leading to myocardial ischemia and reduced myocardial contractility. Myocardial ischemia causes contraction dysfunction, such as contraction delay, weakened contractility, and asynchronous myocardial motion. These changes in cardiac state affect blood flow and are reflected in heart sounds. Consequently, we proposed a noninvasive

Table 13
Ablation experiment results on MDF.

Domain	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
Time	83.797	85.091	81.273	91.561
Cardiac cycle Spectrum	86.456	89.150	81.199	92.914
Entropy	85.051	86.115	86.115	92.813
Energy	83.346	86.229	77.720	90.503
Statistics	84.625	89.112	75.869	91.344
Heart sound spectrum	86.581	91.350	77.276	92.419
Wavelet	86.331	89.568	80.015	92.704

Table 14
Ablation experiment results on MMDF.

Domain	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
Time	84.996	88.720	77.712	90.653
Cardiac cycle Spectrum	85.248	88.264	79.346	91.904
Entropy	85.423	88.416	79.569	92.143
Energy	82.533	86.707	74.368	89.577
Statistics	83.564	81.048	88.484	93.133
Wavelet	85.474	88.606	79.346	92.769

During frequency ablation experiments on wavelet features, the "Base" model integrated with CAIM blocks was used to investigate their contribution to CAD detection across high, medium, and low frequency ranges. The results (as shown in Table 15) indicate that wavelet features in the medium and high frequency ranges contribute more significantly to CAD detection.

Table 15
The results of the frequency ablation experiment on wavelet features.

Domain	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
Wavelet_low	82.910	87.345	74.294	89.135
Wavelet_mid	84.444	84.390	84.547	92.146
Wavelet_high	84.318	83.783	85.364	92.918

Wavelet_low:Wdec_ratio1,Wdec_ratio2,Wdec_l_ratio, Wpac_ratio1,Wpac_l_ratio.
Wavelet_mid:Wdec_ratio3,Wdec_ratio4,Wdec_m_ratio, Wpac_ratio2,Wpac_ratio3,Wpac_ratio4,Wpac_m_ratio; Wavelet_high:Wdec_ratio5,Wdec_h_ratio, Wpac_ratio5,Wpac_ratio6,Wpac_ratio7,Wpac_ratio8,Wpac_h_ratio.

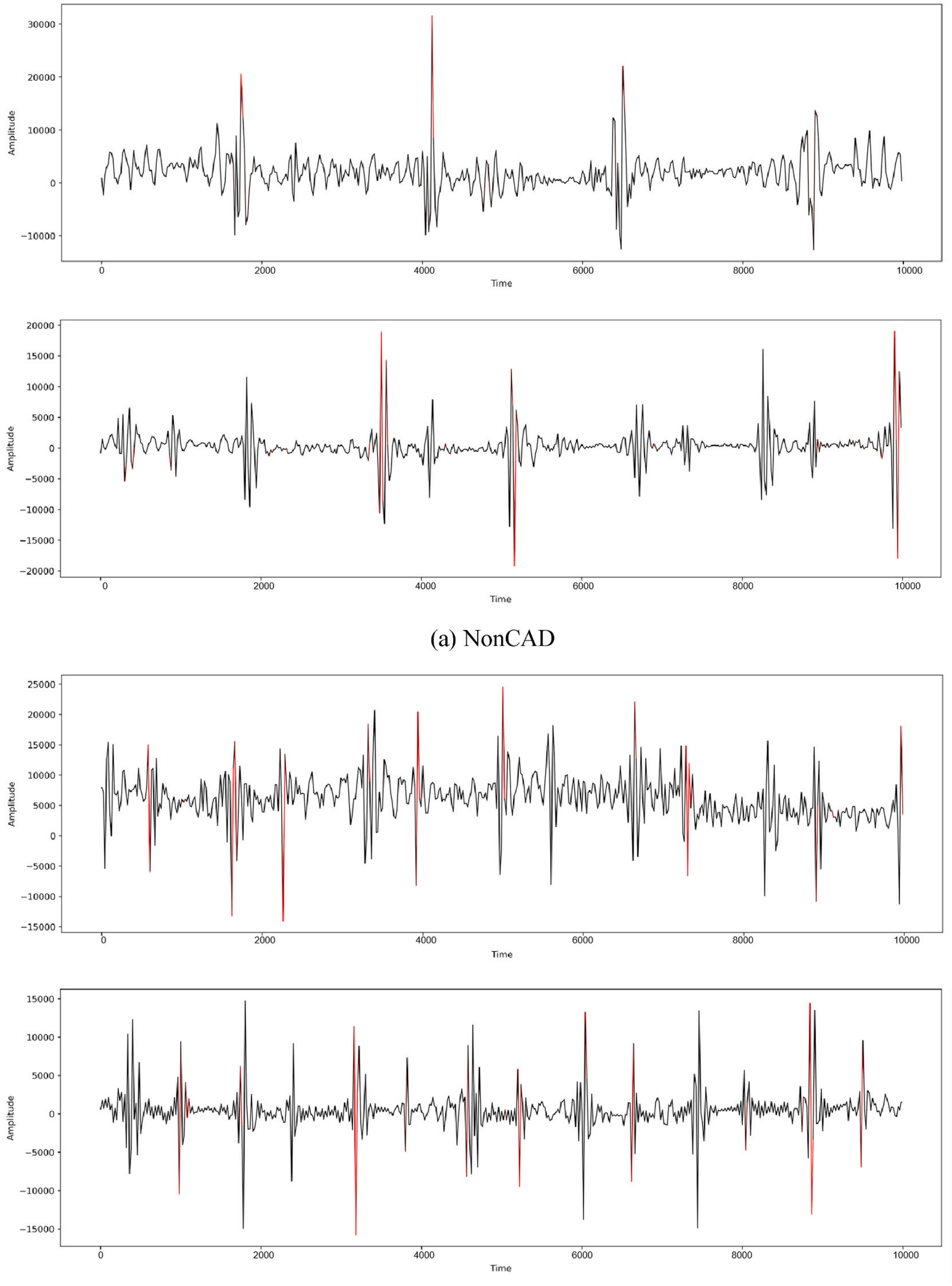
Table 16
Ablation experiment results on MFCC.

Domain	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(%)
Base	83.453	85.628	79.049	90.727
MFCC	86.230	93.171	72.687	91.129

CAD detection method based on heart sounds by utilizing deep learning technologies and a feature fusion framework to facilitate rapid screening and diagnosis of CAD. We designed two models based on data from 400 subjects, the MDF fusion model and the MMDF fusion model, where we achieved AUCs of 94.7 % and 92.7 %, respectively. These two fusion models perform excellently in CAD detection, each with advantages suitable for different needs. The MDF fusion model achieved a higher AUC by integrating more features. In contrast, the MMDF fusion model achieves similar performance with fewer features, with its feature selection aligning better with medical rationale, demonstrating good classification ability and medical adaptability. This method offers an effective and low-cost solution for the early detection and treatment of CAD, particularly suitable for areas with limited medical resources and patients unsuitable for invasive testing.

5.2. Our approach

Traditional heart sound analysis methods typically rely on a few features: time domain [9], frequency domain [20], or time–frequency domain [16]. However, these methods struggle to reveal the complex changes in heart sounds caused by CAD. Time domain features only provide information about the temporal changes in heart sounds, failing to capture crucial aspects such as frequency and energy distribution. Frequency domain analysis reveals the frequency components of heart sounds but does not adequately show their dynamic changes over time and frequency. Time–frequency domain analysis [25] provides rich information but is limited by time and frequency resolution. Traditional methods struggle to precisely capture and interpret complex heart sound characteristics caused by CAD, such as dysfunction and asynchronous motion induced by myocardial ischemia. These characteristics involve multiple frequency ranges and interactions between time and frequency domains, requiring a more comprehensive and detailed analytical approach. To overcome these limitations, we proposed MDF and MMDF to comprehensively capture key information related to CAD detection. Compared to MDF, MMDF combines medical knowledge with MDF, thereby enhancing the reliability and interpretability of CAD detection. Experimental results demonstrated that wavelet features provided ample classification information in the MDF and MMDF fusion models. Frequency ablation experiments revealed that mid-to high-frequency wavelet features significantly contributed to CAD detection, consistent with previous research [52] indicating that partial arterial occlusions cause turbulent blood flow, leading to high-frequency sounds in narrowed areas. Cardiac cycle spectral features also greatly enhanced the performance and reliability of the CAD detection models in both MDF and MMDF. Compared to single-domain features alone, more accurate CAD detection results were obtained using cross-domain features such as wavelet and cardiac cycle spectral features. This finding reveals the limitations of a single-feature domain in diagnosing complex diseases and emphasizes the advantages of employing an MDF fusion strategy to enhance diagnostic precision.



(a) NonCAD

(b) CAD

Fig. 8. Model's interpretability using SHAP algorithms.

5.3. Our model structure

In CAD detection research, heart sound signals are often converted into two-dimensional feature maps for input into models for CAD detection. However, the quality of these two-dimensional feature maps significantly affects the accuracy of CAD detection, as key

Table 17
An overview of the current research on detecting cad through heart sound signals.

Author	Database	Types of Features and Classification Approaches	Result (%)

Akay et al. [49] (2009)	40 subjects: 30 CAD & 10 normal	Approximate entropy
		Optimal threshold detection
			Accuracy = 77.0
			Sensitivity = 78.0
			Specificity = 80.0

Griffel et al. [47] (2012)	31 subjects: 16 CAD & 15 non-CAD	Automutual information function
		Linear support vector machine classifier
			Accuracy = 81.0
			Sensitivity = 87.0
			Specificity = 85.0

Schmidt et al. [50](2015)	133 subjects: 63 CAD & 70 non-CAD	Frequency and nonlinear features
		Quadratic discriminant function
			Accuracy = 68.5
			Sensitivity = 72.0
			Specificity = 65.2

Pathak et al. [51] (2020)	80 subjects: 40 CAD & 40 normal	Imaginary part of cross power spectral density
		Support vector machine classifier
			Accuracy = 75.0
			Sensitivity = 76.5
			Specificity = 73.5

Liu et al. [18](2021)	36 subjects: 21 CAD & 15 normal	Multi-domain and multi-channel features
		Support vector machine classifier
			Accuracy = 90.9
			Sensitivity = 88.0
			Specificity = 93.0

Pathak et al. [48] (2022)	80 subjects:40 CAD & 40 normal	Time-frequency domain and multiple kernel features
		Support vector machine classifier
			Accuracy = 84.19
			Sensitivity = 84.00
			Specificity = 84.38

Ours (MDF)	400 subjects: 268 CAD & 132 normal	Multi-domain and one-dimensional heart Heart sound signal features
		Feature fusion framework
			Accuracy = 87.9
			Sensitivity = 90.7
			Specificity = 82.4

Ours (MMDF)			Accuracy = 85.6
			Sensitivity = 88.0
			Specificity = 80.8

heart sound features are prone to distortion during this conversion process. For instance, Xiang et al. [25] observed significant differences when using different two-dimensional feature maps under the same model. To overcome these issues, we directly used one-dimensional heart sound signals as input and extracted key heart sound features through a one-dimensional convolution model and CAIM. This approach effectively avoided information loss during the conversion process. Simultaneously, CAIM enhanced the ability of the model to extract CAD features. Moreover, we extracted MDF, which was rich in physical meaning, from raw heart sounds. We further extracted MMDF based on medical experience and then fused them to enhance the physical interpretability and detection accuracy of the model. Physical interpretability is particularly important, as it helps doctors better understand and trust the model's predictions. This fusion framework enhances the reliability of the diagnosis and makes the model more efficient and reliable in handling complex heart sound data, enabling more accurate identification and classification of CAD at an early stage, thus providing support for clinical decision making.

5.4. Limitations of our method

Although this study offers an innovative method for noninvasive CAD detection using heart sound analysis, it also has some limitations. The first is the limited number of participants, which constrains the extensive collection of heart sound data and the full training of the model, thereby affecting the enhancement of the model detection's performance. Second is the gold standard for CAD diagnosis: coronary angiography [53], which primarily relies on imaging to identify coronary artery occlusions. However, this method has limited visualization of blood flow details within arteries, especially at specific blockage sites, which may not be easily displayed directly through imaging. Although intravascular ultrasound [54] technology can provide a more comprehensive assessment of vascular and blood flow conditions and offer more accurate CAD detection data, its high cost limits the acquisition of such data. The last is that this study employed one-dimensional convolution modules to extract key features from heart sounds. Although it achieved good detection performance, it did not fully explore more advanced structures that could enhance feature-extraction capability. In addition, the fusion strategy itself has room for further optimization and exploration.

6. Conclusion

This study proposes a fusion framework comprising two CAD detection models, providing patients with a non-invasive detection method. The research collected heart sound data from 400 subjects, using one-dimensional heart sound signals as input to a convolutional neural network to extract heart sound features. Based on the deep learning features of one-dimensional heart sounds, the study fused respectively MDF and MMDF, which are closely related to CAD classification. The combined features were then flattened into low-dimensional features, mapped to binary classification probabilities via a sigmoid function, and used to output CAD detection results. The results indicate that using only a one-dimensional deep learning model for CAD classification achieved an accuracy of 83.5 %, reducing the reliance on empirical dependencies required for spectrum analysis compared to traditional two-dimensional models. After incorporating MDF and MMDF, the MDF fusion model showed accuracies, specificities, sensitivities, and AUCs of 87.9 %, 90.7 %, 82.4 %, and 94.7 %, respectively; the MMDF model had 85.6 %, 88.0 %, 80.8 %, and 92.7 %. The experimental results demonstrate that both models effectively distinguish between CAD and non-CAD cases, enhancing the model's medical interpretability.

Future research can be improved in the following aspects: 1. Expanding the sample size can enhance the generalization ability of the model, helping it perform well on more diverse heart sound signals, thereby improving the accuracy and robustness of CAD detection. 2. Introducing and exploring more advanced deep learning architectures to further enhance the model's feature-extraction capabilities. 3. Refining fusion strategies by integrating new feature fusion methods to better combine information from different features, thereby enhancing the diagnostic performance of the model. 4. Exploring the combination of heart sound signals with other physiological signals (such as electrocardiograms and electronic medical records) to develop multimodal diagnostic models, providing a more comprehensive cardiac health assessment.

These improvements will help further enhance the accuracy and reliability of CAD detection and advance noninvasive CAD diagnostic technology.

Funding

This research was supported by key research and development task of "Xinjiang Uygur autonomous region research on key technologies and optimization strategies for individualized precision diagnosis and treatment system of cardiovascular diseases" (No. 2022B03022-3) and president's fund project of Xinjiang University of political science and law "Research and system construction of coronary artery disease diagnosis based on multi-modal medical data" (No. XZZK2023003).

Ethical statement

The present study complied with the term of the Declaration of Helsinki and was approved by the Institutional Review Board of the First Affiliated Hospital of Xinjiang Medical University (Ethics Approval Number: K202108- 19). Participants gave informed consent to participate in the study before taking part.

Data availability statement

Research data is confidential. Data-sharing requests are required to meet the policies of the hospital and the funder.

CRediT authorship contribution statement

YunFei Dai: Writing – original draft, Software, Formal analysis, Data curation. **PengFei Liu:** Writing – review & editing, Investigation. **WenQing Hou:** Writing – review & editing, Validation, Methodology. **Kaisaierjiang Kadier:** Methodology, Data curation. **ZhengYang Mu:** Writing – review & editing, Methodology. **Zang Lu:** Writing – review & editing, Methodology. **PeiPei Chen:** Writing – review & editing, Methodology. **Xiang Ma:** Writing – review & editing, Methodology, Data curation. **JianGuo Dai:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The authors thank all the technique professionals from the catheterization laboratory, for their assistance in the collection of coronary angiographies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35631.

References

[1] World Health Organization, World health statistics overview 2019: monitoring health for the SDGs, sustainable development goals, World Health Organization. World health statistics 2023: monitoring health for the SDGs, sustainable development goals (who.int) (2019). https://www.who.int/publications/i/item/9789240074323.

[2] Editorial Team of China Cardiovascular Health and Disease Report, Summary of the China cardiovascular health and disease report 2022, Chinese Journal of Circulation 38 (6) (2023) 584–606.

[3] G. Gajanan, S. Samant, C. Hovseth, Y.S. Chatzizisis, Case report: invasive and non-invasive hemodynamic assessment of coronary artery disease: strengths and weaknesses, Frontiers in Cardiovascular Medicine 9 (2022) 885249, https://doi.org/10.3389/fcvm.2022.885249.

[4] G. Barone-Rochette, D. Bruere, N. Mansencal, How to explore coronary artery disease? Archives of cardiovascular diseases 112 (10) (2019) 546–549, https://doi.org/10.1016/j.acvd.2019.05.002.

[5] J. Knuuti, H. Ballo, L.E. Juarez-Orozco, A. Saraste, P. Kolh, A.W.S. Rutjes, W. Wijns, The performance of non-invasive tests to rule-in and rule-out significant coronary artery stenosis in patients with stable angina: a meta-analysis focused on post-test disease probability, Eur. Heart J. 39 (35) (2018) 3322–3330, https://doi.org/10.1093/eurheartj/ehy267.

[6] X. Zhou, X. Guo, Y. Zheng, Y. Zhao, Detection of coronary heart disease based on MFCC characteristics of heart sound, Appl. Acoust. 212 (2023) 109583, https://doi.org/10.1016/j.apacoust.2023.109583.

[7] W. Xu, K. Yu, J. Ye, H. Li, J. Chen, F. Yin, Q. Shu, Automatic pediatric congenital heart disease classification based on heart sound signal, Artif. Intell. Med. 126 (2022) 102257, https://doi.org/10.1016/j.artmed.2022.102257.

[8] N.K. Sawant, S. Patidar, N. Nesaragi, U.R. Acharya, Automated detection of abnormal heart sound signals using Fano-factor constrained tunable quality wavelet transform, Biocybern. Biomed. Eng. 41 (1) (2021) 111–126, https://doi.org/10.1016/j.bbe.2020.12.007.

[9] M.G.M. Milani, P.E. Abas, L.C. De Silva, N.D. Nanayakkara, Abnormal heart sound classification using phonocardiography signals, Smart Health 21 (2021) 100194, https://doi.org/10.1016/j.smhl.2021.100194.

[10] J. Fan, S. Tang, H. Duan, X. Bi, B. Xiao, W. Li, X. Gao, Le-lwtnet: a learnable lifting wavelet convolutional neural network for heart sound abnormality detection, IEEE Trans. Instrum. Meas. 72 (2023) 1–14, https://doi.org/10.1109/TIM.2023.3246513.

[11] Z. Wang, K. Qian, H. Liu, B. Hu, B.W. Schuller, Y. Yamamoto, Exploring interpretable representations for heart sound abnormality detection, Biomed. Signal Process Control 82 (2023) 104569, https://doi.org/10.1016/j.bspc.2023.104569.

[12] J. Liu, H. Wang, Z. Yang, J. Quan, L. Liu, J. Tian, Deep learning-based computer-aided heart sound analysis in children with left-to-right shunt congenital heart disease, Int. J. Cardiol. 348 (2022) 58–64, https://doi.org/10.1016/j.ijcard.2021.12.012.

[13] S. Ismail, B. Ismail, I. Siddiqi, U. Akram, PCG classification through spectrogram using transfer learning, Biomed. Signal Process Control 79 (2023) 104075, https://doi.org/10.1016/j.bspc.2022.104075.

[14] A. Yadav, A. Singh, M.K. Dutta, C.M. Travieso, Machine learning-based classification of cardiac diseases from PCG recorded heart sounds, Neural Comput. Appl. 32 (24) (2020) 17843–17856, https://doi.org/10.1007/s00521-019-04547-5.

[15] S. Aziz, M.U. Khan, M. Alhaisoni, T. Akram, M. Altaf, Phonocardiogram signal processing for automatic diagnosis of congenital heart disorders through fusion of temporal and cepstral features, Sensors 20 (13) (2020) 3790, https://doi.org/10.3390/s20133790.

[16] Z. Abduh, E.A. Nehary, M.A. Wahed, Y.M. Kadah, Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers, Biomed. Signal Process Control 57 (2020) 101788, https://doi.org/10.1016/j.bspc.2019.101788.

[17] H. Li, X. Wang, C. Liu, Q. Zeng, Y. Zheng, X. Chu, C. Karmakar, A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection, Comput. Biol. Med. 120 (2020) 103733, https://doi.org/10.1016/j.compbiomed.2020.103733.

[18] T. Liu, P. Li, Y. Liu, H. Zhang, Y. Li, Y. Jiao, X. Wang, Detection of coronary artery disease using multi-domain feature fusion of multi-channel heart sound signals, Entropy 23 (6) (2021) 642, https://doi.org/10.3390/e23060642.

[19] M.A. Soto-Murillo, J.I. Galvan-Tejada, C.E. Galvan-Tejada, J.M. Celaya-Padilla, H. Luna-Garcia, R. Magallanes-Quintanar, H. Gamboa-Rosales, Automatic evaluation of heart condition according to the sounds emitted and implementing six classification methods, in: Healthcare, vol. 9, MDPI, 2021, March, p. 317, https://doi.org/10.3390/healthcare9030317, 3.

[20] A.D. Levin, A. Ragazzi, S.L. Szot, T. Ning, Extraction and assessment of diagnosis-relevant features for heart murmur classification, Methods 202 (2022) 110–116, https://doi.org/10.1016/j.ymeth.2021.07.002.

[21] X. Cheng, Y. Sun, W. Zhang, Y. Wang, X. Cao, Y. Wang, Application of deep learning in multitemporal remote sensing image classification, Rem. Sens. 15 (15) (2023) 3859, https://doi.org/10.3390/rs15153859.

[22] J. Zhang, H. Liu, K. Yang, X. Hu, R. Liu, R. Stiefelhagen, CMX: cross-modal fusion for RGB-X semantic segmentation with transformers, IEEE Trans. Intell. Transport. Syst. (2023), https://doi.org/10.1109/TITS.2023.3300537.

[23] Y. Tian, S. Wang, E. Li, G. Yang, Z. Liang, M. Tan, MD-YOLO: multi-scale Dense YOLO for small target pest detection, Comput. Electron. Agric. 213 (2023) 108233, https://doi.org/10.1016/j.compag.2023.108233.

[24] X. Chen, X. Guo, Y. Zheng, C. Lv, Heart function grading evaluation based on heart sounds and convolutional neural networks, Physical and Engineering Sciences in Medicine 46 (1) (2023) 279–288, https://doi.org/10.1007/s13246-023-01216-9.

[25] M. Xiang, J. Zang, J. Wang, H. Wang, C. Zhou, R. Bi, C. Xue, Research of heart sound classification using two-dimensional features, Biomed. Signal Process Control 79 (2023) 104190, https://doi.org/10.1016/j.bspc.2022.104190.

[26] W. Chen, Z. Zhou, J. Bao, C. Wang, H. Chen, C. Xu, H. Wu, Classifying heart-sound signals based on CNN trained on MelSpectrum and log-MelSpectrum features, Bioengineering 10 (6) (2023) 645, https://doi.org/10.3390/bioengineering10060645.

[27] M.B. Er, Heart sounds classification using convolutional neural network with 1D-local binary pattern and 1D-local ternary pattern features, Appl. Acoust. 180 (2021) 108152, https://doi.org/10.3390/10.1016/j.apacoust.2021.108152.

[28] X. Bao, Y. Xu, E.N. Kamavuako, The effect of signal duration on the classification of heart sounds: a deep learning approach, Sensors 22 (6) (2022) 2261, 10.3390/10.3390/s22062261.

[29] X. Zhou, X. Guo, Y. Zheng, Y. Zhao, Detection of coronary heart disease based on MFCC characteristics of heart sound, Appl. Acoust. 212 (2023) 109583, https://doi.org/10.3390/sound/10.1016/j.apacoust.2023.109583.

[30] D.B. Springer, L. Tarassenko, G.D. Clifford, Logistic regression-HSMM-based heart sound segmentation, IEEE Trans. Biomed. Eng. 63 (4) (2015) 822–832, https://doi.org/10.1109/TBME.2015.2475278.

[31] H.J. Kim, I.E. Vignon-Clementel, J.S. Coogan, C.A. Figueroa, K.E. Jansen, C.A. Taylor, Patient-specific modeling of blood flow and pressure in human coronary arteries, Ann. Biomed. Eng. 38 (2010) 3195–3209, https://doi.org/10.1007/s10439-010-0083-6.

[32] C. Liu, D. Springer, Q. Li, B. Moody, R.A. Juan, F.J. Chorro, G.D. Clifford, An open access database for the evaluation of heart sound algorithms, Physiol. Meas. 37 (12) (2016) 2181, https://doi.org/10.1088/0967-3334/37/12/2181.

[33] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (6) (2000) H2039–H2049, https://doi.org/10.1152/ajpheart.2000.278.6.H2039.

[34] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy, IEEE Trans. Neural Syst. Rehabil. Eng. 15 (2) (2007) 266–272, https://doi.org/10.1109/TNSRE.2007.897025.

[35] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, Y. Hou, Assessing the complexity of short-term heartbeat interval series by distribution entropy, Med. Biol. Eng. Comput. 53 (2015) 77–87, https://doi.org/10.1007/s11517-014-1216-0.

[36] H. Tang, Z. Dai, Y. Jiang, T. Li, C. Liu, PCG classification using multidomain features and SVM classifier, BioMed Res. Int. 2018 (2018), https://doi.org/10.1155/2018/4205027.

[37] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693, https://doi.org/10.1109/34.192463.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.

[39] 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, 10.1109/10.1109/CVPR.2016.90.

[40] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474.

[41] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11534–11542, https://doi.org/10.1109/CVPR42600.2020.01155.

[42] A. Ainiwaer, W.Q. Hou, Q. Qi, K. Kadier, L. Qin, R. Rehemuding, Y.T. Ma, Deep learning of heart-sound signals for efficient prediction of obstructive coronary artery disease, Heliyon 10 (1) (2024) e23354, https://doi.org/10.1016/j.heliyon.2023.e23354.

[43] A. Srinivas, T.Y. Lin, N. Parmar, J. Shlens, P. Abbeel, A. Vaswani, Bottleneck transformers for visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16519–16529, https://doi.org/10.1109/CVPR46437.2021.01625.

[44] Y. Li, T. Yao, Y. Pan, T. Mei, Contextual transformer networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2) (2022) 1489–1500, https://doi.org/10.1109/TPAMI.2022.3164083.

[45] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, Proc. AAAI Conf. Artif. Intell. 31 (1) (2017, February), https://doi.org/10.1609/aaai.v31i1.11231.

[46] A. Howard, M. Sandler, G. Chu, L.C. Chen, B. Chen, M. Tan, H. Adam, Searching for mobilenetv3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314–1324, https://doi.org/10.1109/ICCV.2019.00140.

[47] M. Akay, Y.M. Akay, D. Gauthier, R.G. Paden, W. Pavlicek, F.D. Fortuin, R.W. Lee, Dynamics of diastolic sounds caused by partially occluded coronary arteries, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 56 (2) (2008) 513–517, https://doi.org/10.1109/TBME.2008.2003098.

[48] A. Pathak, K. Mandana, G. Saha, Ensembled transfer learning and multiple kernel learning for phonocardiogram based atherosclerotic coronary artery disease detection, IEEE Journal of Biomedical and Health Informatics 26 (6) (2022) 2804–2813, https://doi.org/10.1109/JBHI.2022.3140277.

[49] B. Griffel, M.K. Zia, V. Fridman, C. Saponieri, J.L. Semmlow, Microphone placement evaluation for acoustic detection of coronary artery disease, in: 2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC), IEEE, 2011, April, pp. 1–2, https://doi.org/10.1109/NEBC.2011.5778616.

[50] S.E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, J.J. Struijk, Acoustic features for the identification of coronary artery disease, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 62 (11) (2015) 2611–2619, https://doi.org/10.1109/TBME.2015.2432129.

[51] A. Pathak, P. Samanta, K. Mandana, G. Saha, An improved method to detect coronary artery disease using phonocardiogram signals in noisy environment, Appl. Acoust. 164 (2020) 107242, https://doi.org/10.1016/j.apacoust.2020.107242.

[52] J. Semmlow, W. Welkowitz, J. Kostis, J.W. Mackenzie, Coronary artery disease-correlates between diastolic auditory characteristics and coronary artery stenoses, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. (2) (1983) 136–139, https://doi.org/10.1016/10.1109/TBME.1983.325211.

[53] D. Giacoppo, C. Laudani, G. Occhipinti, M. Spagnolo, A. Greco, C. Rochira, D. Capodanno, Coronary angiography, intravascular ultrasound, and optical coherence tomography for guiding of percutaneous coronary intervention: a systematic review and network meta-analysis, Circulation 149 (14) (2024) 1065–1086, https://doi.org/10.1161/CIRCULATIONAHA.123.067583.

[54] X. Li, Z. Ge, J. Kan, M. Anjum, P. Xie, X. Chen, L. Hong, Intravascular ultrasound-guided versus angiography-guided percutaneous coronary intervention in acute coronary syndromes (IVUS-ACS): a two-stage, multicentre, randomised trial, Lancet (2024), https://doi.org/10.1016/S0140-6736(24)00282-4.

Abbreviation of this study

CAD: Coronary Artery Disease
CCTA: Coronary Computed Tomography Angiography
MFCC: Mel-Frequency Cepstral Coefficients
MDF: Multi-Domain Feature
MMDF: Medical Multi-Domain Feature
CNN: Convolutional Neural Networks
MLP: Multilayer Perceptron
CAIM: Channel Attention Inversion Module
BN: Batch Normalization
ECA: Efficient Channel Attention