International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 7904-7925

https://doi.org/10.70517/ijhsa464679

Development of a physiological signal processing and prediction model for heart failure patients

Jinlong Zhuang[1], Taoming Qian[1] and Li Liu[2,*]
[1] Graduate School, Heilongjiang University of Traditional Chinese Medicine, Harbin, Heilongjiang, 150040, China
[2] The First Affiliated Hospital, Heilongjiang University of Traditional Chinese Medicine, Harbin, Heilongjiang, 150040, China
Corresponding authors: (e-mail: valen333@126.com).

Abstract Cardiovascular disease is the leading cause of death in humans, and heart failure is the primary cause of death among cardiovascular diseases, significantly impacting patients' quality of life and life expectancy. This paper proposes corresponding heart sound signal analysis methods based on the physiological structure of the cardiovascular circulatory system, combining three aspects: cardiac reserve indicators, energy characteristics, and complexity characteristics. Subsequently, experimental studies were conducted to investigate the differences in short-term heart sound features between chronic heart failure patients and healthy individuals, as well as the relationship between short-term heart sound features and the staging of chronic heart failure. Finally, the MSCNN-MGU heart failure prediction model was established by combining MSCNN with MGU. The results of this study indicate that as the severity of heart failure increases, the D/S and S1/S2 ratios in the time-domain features of heart sounds show a decreasing trend. In performance comparison experiments among different neural network models, the performance of EfficientNet-B2 was as follows: Acc=94.13, Pre=92.38, Rec=81.93, F1=84.5, AUC=0.945, and the inference speed with BatchSize set to 128 was 917ms. Thus, the model achieves high balanced performance while ensuring fast inference speed.

Index Terms heart failure, heart sound signals, MSCNN-MGU heart failure prediction model, short-term features of heart sounds

I. Introduction

In recent years, with the rapid and healthy development of the economy, people's material living standards have significantly improved. However, unhealthy lifestyles have become increasingly prevalent, and issues such as an aging population have contributed to a continuous rise in the prevalence of cardiovascular diseases [1]-[3]. Heart failure represents the severe and terminal stage of various chronic heart diseases, posing a higher risk compared to other cardiovascular conditions, and significantly impacting patients' quality of life and life expectancy [4], [5]. Common symptoms of heart failure include fatigue, shortness of breath, reduced exercise tolerance, and fluid retention, which can further lead to congestion in the internal organs and lungs [6], [7]. Therefore, researching diagnostic techniques for early-stage heart failure can effectively reduce patient mortality and hospitalization rates, and timely screening for heart failure is of great significance for disease prevention and improving prognosis [8], [9].

Currently, the diagnosis of heart failure is primarily based on a comprehensive assessment of physical examination, cardiac imaging studies, and other auxiliary tests. However, these methods are hindered by high costs, complex testing processes, inconvenient operation, observer bias, and invasive procedures, making them unsuitable for widespread adoption and routine use [10]-[13]. Non-invasive automatic disease diagnosis technology based on physiological signal detection and analysis is one of the primary research topics in the field of biomedical engineering [14]. For early diagnosis of heart failure, it is possible to study the electrophysiological characteristics of the heart under different states, particularly through non-invasive detection of physiological signals and surface features that reflect cardiac pathology [15]-[17]. By leveraging rapidly advancing computer-aided diagnostic technologies and integrating existing artificial intelligence methods, an automated diagnostic model for heart failure can be established to assist physicians in clinical diagnosis [18]-[20].

Heart sounds and electrocardiogram (ECG) signals reflect the electrical and mechanical activity characteristics of the heart, both of which can be obtained using non-invasive methods and are easy to examine medically. Therefore, analysis methods based on these two signals provide possibilities for automated detection of heart failure. On one hand, some scholars have conducted research on heart failure detection based on heart sound signals. Literature [21] points out that heart sound signals recorded by electronic stethoscopes carry useful information about cardiac function and status. Therefore, using machine learning techniques to extract features and classify heart sound signals can significantly improve the efficiency of early diagnosis of heart diseases such as heart failure

(CHF). Reference [22] created a mobile health service platform for analyzing and classifying heart sound data, enabling CHF patients to undergo testing at low cost and share analysis results with healthcare providers via the internet, thereby achieving remote personalized diagnosis. Reference [23] combined machine learning and end-to-end deep learning techniques to construct an automatic monitoring model for the diagnosis of chronic heart failure. By learning and training on expert features in the heart sound dataset, the model demonstrated good performance in both the diagnosis and staging of chronic heart failure. Literature [24] investigates the design of heart sound classification algorithms, employing two segmentation criteria in heart sound analysis to detect cycles with minimal noise interference and optimal segmentation results. These cycles are then selected to extract time-domain, time-frequency-domain, and perceptual-domain features from the variability of heart sound signals, thereby mitigating noise interference in heart sound analysis. Literature [25] also proposed a machine learning method for detecting chronic heart failure from heart sound signals, applying it to the clinical diagnosis of CHF patients, demonstrating high detection accuracy, and providing an important boost to improving patients' quality of life.

On the other hand, scholars have conducted research on heart failure detection based on electrocardiogram (ECG) signals. Literature [26] highlights that ECG signal analysis possesses cost-effective and widespread characteristics not found in other diagnostic methods. However, since changes in ECG signals are typically not specific to CHF diagnosis, a computer-aided detection system is required to enhance diagnostic objectivity. Literature [27] indicates that computer-aided diagnostic systems for ECG signals can provide more objective and faster interpretation of ECG signals. Therefore, a deep convolutional neural network-based automatic diagnostic model for CHF was proposed to assist cardiologists in diagnostic and therapeutic activities. Literature [28] analyzes the time-frequency subband matrix of ECG signals using Stockwell transform and frequency division techniques, and designs an automatic detection strategy for CHF based on this analysis, demonstrating high accuracy, sensitivity, and specificity in the field of congestive heart failure disease detection. Literature [29] addresses the issues of similarity in time intervals and amplitude of ECG signals, as well as noise problems, by proposing the use of an improved support vector machine to analyze ECG signals for identifying heart failure diseases, thereby supporting the treatment and care of patients with heart diseases. Literature [30] introduces a method for assessing cardiac structural and functional abnormalities using echocardiography measurement results, and incorporates knowledge-enhanced neural networks to process biological signals during diagnosis, effectively reducing diagnostic gaps in care and improving the predictive performance of heart failure disease.

The above studies have confirmed the important clinical research value of ECG and heart sound signals in non-invasive and non-destructive detection of heart failure. However, the studies are somewhat lacking in terms of signal data processing, and the computational capabilities of detection models and disease prediction accuracy still need to be further improved.

The article first introduces the physiological knowledge of the cardiovascular system, including the human blood circulation process, the structure of the heart, and the mechanism of heart sound generation. It then proposes a shoulder-mounted heart sound stethoscope and common heart sound signal denoising methods. Subsequently, through time-domain analysis, energy analysis, and complexity analysis of heart sounds, both linear and nonlinear features are extracted, including time-domain cardiac reserve indicators, subband energy fraction, component energy distribution characteristics, and the complexity of heart sound time series, totaling 13 feature indicators. The study also investigates the differences in short-term heart sound features between patients with chronic heart failure and healthy individuals, as well as the relationship between short-term heart sound features and the staging of chronic heart failure. Finally, the paper combines MSCNN with MGU to establish an MSCNN-MGU heart failure classification model, and validates the effectiveness of its structural design through performance testing experiments.

II. Preprocessing of cardiovascular system and heart sound signals

II. A. Cardiovascular System

The cardiovascular system is a tubular system that can be divided into two components: the heart and blood vessels. Blood flows through the blood vessels, with the heart acting as a pumping organ to provide the force for blood circulation. Blood vessels can be further classified into arteries, veins, and capillaries. The process by which the heart's rhythmic contractions drive blood to circulate in a directed manner through the blood vessels is known as blood circulation. Blood circulation can be divided into two parts based on the organs involved: systemic circulation and pulmonary circulation. Systemic circulation begins in the left ventricle. Blood from the left ventricle flows through the aortic valve into the aorta, then travels through the arterial system to the capillaries in various parts of the body, where it enters the veins. The blood then returns to the right atrium via the superior vena cava and inferior vena cava, flows into the right ventricle, and completes the circulation process. The blood vessels involved in pulmonary circulation include the pulmonary artery and pulmonary veins. Blood from the right ventricle flows out through the pulmonary valve into the pulmonary arteries, directly reaching the pulmonary capillaries. In the pulmonary capillaries, gas exchange occurs with the gases in the alveoli, at which point the blood becomes bright red arterial blood. It then flows back to the left atrium via the pulmonary veins, completing the pulmonary circulation process. The transportation of substances within the body primarily relies on the circulatory system. If blood circulation ceases, the body's metabolism will become abnormal.

II. B.Heart Structure

The position of the heart within the human body changes with breathing, body posture, and position. Typically, in adults, the right half of the heart is located in the anterior superior region, while the left half is in the posterior inferior region. The heart has a shape resembling a cone. Its size is generally comparable to the size of one's fist, with a length of approximately 12–14 cm, a transverse diameter of approximately 9–11 cm, and an anteroposterior diameter of approximately 6–7 cm. The weight of the heart typically accounts for 0.5% of body weight, averaging 260 g. The right atrium is located in the upper right portion of the heart, with thin walls and a large cavity. The portion that protrudes toward the left front is referred to as the right atrial appendage.

II. C.Mechanism of heart sound generation

Heart sounds follow a certain pattern and exhibit approximate periodic characteristics. A cardiac cycle consists of two phases: systole and diastole. The cardiovascular blood circulation diagram is shown in Figure 1. The primary driving force of the heart comes from the periodic relaxation and contraction of the myocardium. The contraction and relaxation of the myocardium, in coordination with the opening and closing of the valves, cause the pressure and volume of the atria and ventricles to change rhythmically, thereby propelling blood to circulate throughout the body. The cardiac cycle begins with ventricular contraction. During a complete cardiac cycle, four heart sounds are produced. These are named in the order of their appearance in the cardiac cycle: first heart sound (S1), second heart sound (S2), third heart sound (S3), and fourth heart sound (S4). Under normal circumstances, S1 and S2 can be heard through a stethoscope. S1 occurs during ventricular systole, lasting approximately 0.1 seconds, with a frequency of 40–60 Hz, marking the onset of ventricular contraction. The contraction of the left and right ventricular muscles, along with the closure of the mitral and tricuspid valves, generates a series of mechanical signals. As the aortic and pulmonary valves open and blood is ejected into the aorta, this causes vibrations in the ventricular walls and the walls of the great arteries. These mechanical signals constitute the primary components of S1. S1 can be clearly heard at the apex of the heart. S2 occurs at the beginning of the myocardial diastole, lasting approximately 0.08 seconds, with a frequency typically ranging from 60 to 100 Hz. The rapid closure of the aortic and pulmonary valves, the deceleration of blood flow in the large arteries, and the rapid decrease in intraventricular pressure cause vibrations that constitute the primary component of S2. S2 can be clearly heard in the aortic and pulmonary artery auscultation areas. S3 occurs at the beginning of ventricular diastole and the end of rapid filling. It is believed to be caused by the impact of blood rapidly entering the left or right ventricle from the left or right atrium, causing vibrations in the ventricular wall. Since S3 is primarily caused by changes in blood flow velocity, its average frequency is relatively low, approximately 10–20 Hz, with a duration of about 0.04 seconds. S4 occurs at the end of ventricular diastole and before systole, resulting from the vibration caused by the atrium muscle contracting forcefully at the end of systole to overcome the pressure at the end of ventricular diastole. It has a low frequency and lasts approximately 0.04 seconds. For a beating heart, these repetitive mechanical vibration signals transmitted through the thoracic conduction system to the body surface form the periodic “heart sounds” audible to the human ear.

II. D.Acquisition of heart sound signals

To collect heart sounds over an extended period of time, the shoulder strap-type stethoscope invented by our research team is used. This device consists of a lightweight elastic material forming an Ω -shaped frame assembly, whose shape resembles the lateral contour line from the human shoulder to the chest, allowing the frame to be conveniently placed on the left shoulder. The length of both ends of the Ω -shaped frame extends from the top of the left shoulder to the center of the human heart. A heart sound sensor is installed at one end of the Ω shaped frame, and the length of this end can be partially adjusted to ensure that the heart sound sensor remains accurately positioned at the apex of the heart. The other end of the Ω shaped frame can symmetrically install another heart sound sensor, enabling one sensor to collect heart sounds from the chest position and the other from the back position, thereby obtaining two channels of stereo heart sounds.

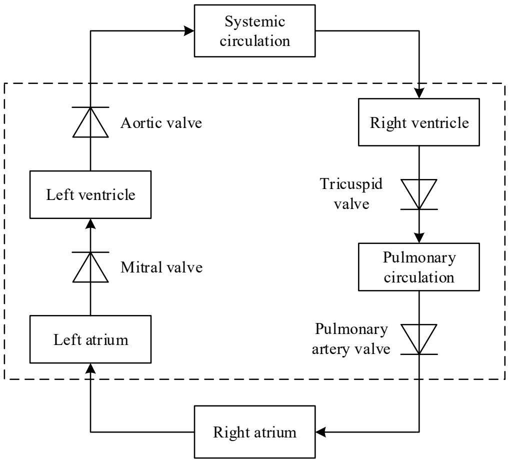


Figure 1: Cardiovascular blood circulation block diagram

II. E. Heart sound signal denoising

Characteristics of Wavelet Denoising The wavelet transform inherits the characteristics of the Fourier transform and further develops and improves upon them. The multi-resolution analysis of wavelets exhibits excellent spatial and frequency domain localization properties, enabling it to focus on any detail of the signal being analyzed. As a result, it is particularly well-suited for processing non-stationary signals such as heart sounds. Due to the translation and scaling capabilities of wavelet bases, wavelets possess a flexible and adjustable time-frequency window. At high frequencies, the time window narrows, while at low frequencies, it widens, facilitating the resolution of non-stationary heart sound signals.

In the field of wavelet analysis, $L^2(R)$ is the most frequently discussed function space. $L^2(R)$ denotes the function space consisting of square-integrable functions on R, i.e., $f(t) \in L^2(R) \Leftrightarrow \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty$. Exploring how to represent $L^2(R)$ functions using the binary contraction and translation of a basic wavelet function is one of the most important problems in wavelet analysis [31].

There are three main forms of wavelet transforms: continuous wavelet transform, discrete wavelet transform, and binary wavelet transform.

(1) Continuous wavelet transform

Let $\varphi(t)$ be the basic wavelet. By scaling and translating the basic wavelet, we obtain:

$$\varphi_{a,b} = \frac{1}{\sqrt{|a|}} \varphi\left(\frac{t-b}{a}\right) \quad a,b \in R, a \neq 0 \tag{1}$$

For $f(t) \in L^2(R)$, its continuous wavelet transform can be defined as:

$$WT_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t)\varphi^*\left(\frac{t-b}{a}\right) dt = \left\langle f, \varphi_{a,b} \right\rangle \tag{2}$$

In the equation, a is the scale factor. b is the translation factor. $\varphi^*(t)$ is the complex conjugate of $\varphi(t)$.

(2) Discrete wavelet transform

Compared to the continuous wavelet transform, let the parameters $a = 2^{-j}$ and $b = k \cdot 2^{-j}$, where $j,k \in Z$, then the discrete wavelet can be expressed as:

$$\varphi_{2^{-j},k\cdot 2^{-j}} = 2^{j/2}\varphi(2^j t - k) \tag{3}$$

Typically, $\varphi_{j,k}(t)$ is used to represent $\varphi_{2^{-j},k\cdot 2^{-k}}$, and the discrete wavelet transform is defined as:

$$WT_f(j,k) = \left\langle f, \varphi_{j,k} \right\rangle = 2^{j/2} \int_{-\infty}^{+\infty} f(t)\varphi^*(2^j t - k) dt \tag{4}$$

(3) Binary wavelet transform

Compared with the continuous wavelet transform, let the parameter $a = 2^{-j}$, $j \in Z$, and the parameter b still take continuous values, then we obtain the binary wavelet:

$$\varphi_{2^j,b}(t) = 2^{-j/2} \varphi\left[2^{-j}(t-b)\right] \tag{5}$$

At this point, the binary wavelet transform of $f(t) \in L^2(R)$ can be defined as:

$$WT_f(2^j, b) = 2^{-j/2} \int_{-\infty}^{+\infty} f(t)\varphi^*[2^{-j}(t-b)]dt \tag{6}$$

From the above definition, it can be seen that the binary wavelet transform discretizes the scale parameter while maintaining continuous variation in the shift amount. Clearly, the binary wavelet lies between continuous wavelets and discrete wavelets. Unlike the discrete wavelet transform, the binary wavelet transform retains the advantage of shift invariance characteristic of the continuous wavelet transform.

Given the characteristics of cardiac sound signals, selecting an appropriate wavelet basis function is particularly important to ensure the effectiveness, accuracy, and operational efficiency of the signal processing process. Typically, the application of wavelet bases involves using a small number of non-zero wavelet coefficients to adequately approximate a specific class of functions. By optimizing the design of the wavelet φ, the maximum number of wavelet coefficients $\left\langle f, \varphi_{j,k} \right\rangle$ close to zero can be generated. If most of the small-scale wavelet coefficients are small, then the number of wavelet coefficients in the function f that cannot be ignored will be small. This characteristic of wavelet analysis is primarily determined by the regularity of f, the vanishing moment order of φ, and the support size of φ.

III. Heart sound feature extraction

III. A. Time Domain Feature Extraction

III. A. 1) Concept of cardiac reserve

Cardiac reserve is related to cardiac upregulation capacity and reflects the maximum capacity available to the body under stress. Cardiac reserve (CR) indicators belong to the time-domain characteristics of heart sounds and mainly include the following two items:

(1) D/S

D/S is defined as the ratio of the duration of diastole to the duration of systole (D/S) during the cardiac cycle. During diastole, the heart is in a gradual filling phase, which is related to its own blood supply. The duration of diastole, i.e., the length of time blood perfusion occurs, affects the heart's metabolic environment. A healthy heart has a longer diastolic phase duration, providing sufficient blood supply time to deliver adequate oxygen and nutrients during the systolic phase, which is beneficial for the heart's metabolic environment and maintains it in a healthy state with normal myocardial contractility. Conversely, if the diastolic phase duration is too short, it may result in insufficient blood perfusion time in the ventricles, inadequate self-blood supply, and an inability to provide sufficient nutrients and oxygen, thereby impairing myocardial contractility. D/S serves as an indicator for assessing cardiac reserve and can be used to determine whether the heart's self-perfusion time during the cardiac cycle is adequate. The D/S value is crucial for ensuring the healthy functioning of the heart's pumping capacity.

(2) S1/S2

S1/S2 is the ratio of the amplitude of the first heart sound to the second heart sound (S1/S2). The first heart sound directly reflects myocardial contractility. The second heart sound reflects the magnitude of peripheral circulatory resistance during cardiac pumping. Due to significant individual variability in heart sounds, the absolute value has limited practical application value. Therefore, relative values are used to improve the accuracy of cardiac assessment [32]. The S1/S2 ratio reflects the dynamic relationship between cardiac contractility and peripheral resistance, serving as a comprehensive indicator for assessing myocardial contractility and peripheral circulatory resistance.

III. A. 2) Extraction of cardiac reserve indicators

Based on the segmentation of heart sounds, the segmentation points of each heart sound component are obtained. In array B, the starting points of a series of alternating $S1$ and $S2$ are recorded. After determining the order of appearance, assume that $B[1]$ and $B[1+2n]$ both represent the starting point of the first heart sound, $B[2]$ and $B[2+2n]$ represent the starting points of the second heart sound. Similarly, array C contains the ending points of

$S1$ and $S2$, where $C\left[1\right]$ and $C\left[1+2n\right]$ both represent the ending points of the first heart sound, $C\left[2\right]$ and $C\left[2+2n\right]$ both represent the end point of the second heart sound.

(1) D/SThe ratio of the diastolic period to the systolic period is represented by D/S, which is calculated as follows:

Diastolic period:

$$T_d = B[1+2n] - B[2n] \tag{7}$$

Systolic duration:

$$T_s = B[2n] - B[2n-1] \tag{8}$$

The value of the time domain characteristic D/S of the heart sound signal can be calculated using the following formula:

$$D/S = \frac{T_d}{T_s} \tag{9}$$

(2) $S1/S2$

The amplitude ratio of the first heart sound or second heart sound amplitude $S1/S2$ is calculated as follows:

Based on the heart sound envelope, all peak points are located using a differential algorithm, which includes the first heart sound or second heart sound amplitude, but there may be false peaks that cause interference with $S1$ and $S2$. The method for removing false peaks involves combining the localization information of the heart sound signal with the start and end points of $S1$ and $C\left[1+2n\right]:B\left[1+2n\right]$ and $S2$ start and end points $C\left[2+2n\right]:B\left[2+2n\right]$ as constraints, and taking the maximum peak value within this range as the amplitude of the current heart sound component.

III. B. Energy Feature Extraction

III. B. 1) Frequency domain energy distribution characteristics

Based on the frequency domain characteristics of heart sounds, it is known that the primary components of normal heart sounds, S1 and S2, are distributed within a specific frequency range, with S1 occurring at a slightly lower frequency than S2. Additionally, the distribution of heart sound energy in the frequency domain follows certain patterns. In patients with heart failure, impaired cardiac function and hemodynamic abnormalities can affect heart sounds, leading to changes in their frequency distribution. By analyzing the energy distribution of heart sounds in the frequency domain, abnormalities in each heart sound component can be identified. Commonly used methods for analyzing the frequency domain energy distribution of heart sounds include the fast Fourier transform, wavelet decomposition, and wavelet packet decomposition. In this study, wavelet packet decomposition was selected for frequency domain energy analysis of heart sound signals, and their frequency domain energy distribution characteristics were extracted.

The mathematical description of wavelet packets is as follows:

$$\varphi_{j,k}^{i}(t) = 2^{-\frac{j}{2}} \varphi_n^i\left(2^{-j}t - k\right) \tag{10}$$

It is composed of a set of linearly combined wavelet functions and can be represented by the following recursive relationship:

$$\begin{cases} \varphi^{2i} = \dfrac{1}{\sqrt{2}} \sum\limits_{-\infty}^{\infty} h(k)\varphi^i\left(\dfrac{t}{2} - k\right) \\ \varphi^{2i+1} = \dfrac{1}{\sqrt{2}} \sum\limits_{-\infty}^{\infty} g(k)\varphi^i\left(\dfrac{t}{2} - k\right) \end{cases} \tag{11}$$

Among these, $\varphi^1(t) = \varphi(t)$ is the wavelet function, $\varphi^0(t) = \varnothing(t)$ is the scale function, and $h(k)$ and $g(k)$ represent a pair of conjugate mirror filter coefficients associated with the scale function and wavelet function in wavelet multiscale analysis, and they satisfy the formula $g(k) = (-1)^k h(1-k)$. The wavelet packet decomposition coefficients

$c_{j,k}^{i}$ of the signal $f(t)$ are calculated using the following formula under the condition that the wavelet coefficients satisfy the orthogonality condition:

$$c_{j,k}^{i}=\int_{-\infty}^{\infty} f(t)\varphi_{j,k}^{i}(t)dt \tag{12}$$

The signal $f(t)$ can be expressed by the following formula after being decomposed by the wavelet packet i layer:

$$\begin{aligned} f(t) &= \sum_{j=0}^{2^i-1} f_{i,j}\left(t_j\right) \\ &= f_{i,0}\left(t_0\right)+f_{i,1}\left(t_1\right)+L+f_{i,2^i-1}\left(t_{2^i-1}\right), j=0,1,2,L,2^i-1 \end{aligned} \tag{13}$$

The energy at each node after signal decomposition can be calculated using the following formula:

$$E_{i,j}\left(t_j\right)=\int_{t_j}\left|f_{i,j}\left(t_j\right)\right|^2 dt=\sum_{k=1}^{m}\left|x_{j,k}\right|^2 \tag{14}$$

The band energy at node (i,j) is $E_{i,j}\left(t_j\right)$, and $f_{i,j}\left(t_j\right)$ is the reconstructed signal, whose discrete point amplitude is $x_{j,k}\left(j=0,1,2,L,2^i-1;k=1,2,L,m\right)$, where m is the number of sampling points of the signal.

III. B. 2) Cardiac energy efficiency index

From the process of cardiac energy utilization, it is known that the energy stored in the heart is primarily converted into pressure energy through myocardial contraction to pump blood outward. Myocardial contraction occurs during the systolic phase of the heart, during which left ventricular pressure increases. The amplitude of S1 is positively correlated with the rate of increase in left ventricular pressure. Therefore, the magnitude of pressure energy, which represents the energy expenditure used for pumping blood, can be measured by the energy of S1, corresponding to the useful work performed by cardiac energy utilization. This section introduces the concept of cardiac energy efficiency, which can be reflected by the energy fraction of heart sounds to indicate whether there are energy utilization disorders or energy waste in the heart. Cardiac energy efficiency indicators include the “first heart sound energy fraction” and the “heart murmur energy fraction.”

The following are the calculation formulas:

First heart sound energy fraction:

$$S1\_EF=\frac{E(S1)}{E(heartsound)}\times 100\% \tag{15}$$

Noise energy fraction:

$$HM\_EF=\frac{E(murmur)}{E(heartsound)}\times 100\% \tag{16}$$

Among them,

$$E(heartsound)=E(S1)+E(S2)+E(S3)+E(S4)+E(murmur) \tag{17}$$

Energy formula for each component of heart sounds:

$$E=\sum_{i=n}^{n+N}\left|x(i)\right|^2 \tag{18}$$

$x(i)$ is the discrete sampling sequence of heart sound samples, and N denotes the number of sampling points.

III. C. Heart sound complexity feature extraction

Sample entropy is an improvement on approximate entropy. Sample entropy values can be used to characterize the complexity of signal time series at different scales; the larger the sample entropy value, the more complex the sequence. It is also a measure of the rate of new information generation in signals and can be used to evaluate the adaptive variability of a system from the perspective of time series complexity.

The calculation steps for sample entropy are as follows:

(1) Reconstruct the original sequence $\{X_t\} = \{x_1, x_2, \ldots x_N\}$ into an m-dimensional vector according to the given embedding dimension m, that is:

$$X(i) = [x_i, x_{i+1}, \ldots, x_{i+m-1}], \text{ Which } i = 1 \sim N - m + 1 \tag{19}$$

(2) Define the distance d_{ij} between $x(i)$ and $x(j)$ as the largest absolute value of the difference between their corresponding elements, i.e.:

$$d_{ij} = d[x(i), x(j)] = \max\left[\left|x(i+k) - x(j+k)\right|\right], k = 0 \sim m-1 \tag{20}$$

(3) Given a threshold r, calculate the distance d_{ij} between $x(i)$ and the remaining vectors $x(j)(j = 1, 2, \ldots, N-m; j \neq i)$ for each value of i, and calculate the ratio of the number of $d_{ij} < r$ to the total number of vectors, denoted as:

$$B_i^m(r) = \frac{1}{N-m} \text{num}\{dij < r\}, i = 1 \sim N - m + 1, i \neq j \tag{21}$$

(4) Find the average value of all $B_i^m(r)$:

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \tag{22}$$

Increase the dimension by 1, repeat steps ① to ④ for the $m+1$ dimension, and further obtain $B_i^m(r)$ and $B^{m+1}(r)$.

(5) Calculate the sample entropy of the original sequence:

$$SampEn(m, r) = \lim_{N \to \infty} \left[-\ln \frac{B^{m+1}(r)}{B^m(r)} \right] \tag{23}$$

In practical applications, N is generally a finite number, in which case the sample entropy is expressed as:

$$SampEn(m, r, N) = \ln B^m(r) - \ln B^{m+1}(r) \tag{24}$$

Sample entropy is chosen to measure the nonlinear characteristics of heart sound signals mainly because it has the following characteristics:

(1) It avoids errors caused by comparison with its own data segments, and the data length has little effect on the sample entropy value, resulting in accurate estimates.

(2) It exhibits better relative consistency, meaning that sample sequences with higher SampEn values maintain high SampEn values under different embedding dimensions m and similarity thresholds r.

(3) It is applicable to both random and deterministic processes, making it particularly suitable for quantifying the nonlinearity and non-stationarity of systems.

It is insensitive to missing data, with only minor effects on SampEn values from small amounts of data loss.

IV. Physiological signal data feature extraction based on electrocardiogram signals

IV. A. ECG Data Processing

IV. A. 1) Electrocardiogram Dataset

All data used in this study were obtained from the PhysioNet database of complex physiological signals. The data for the normal group were sourced from the MIT-BIH Normal Sinus Rhythm Database, which includes long-term electrocardiogram (ECG) recordings from 18 patients. The participants included in this database had no significant arrhythmias, comprising 10 males (aged 26 to 45 years) and 8 females (ages 20 to 50). The average recording duration for each case was approximately 22 hours, including ECG data from two channels, with a sampling frequency of 128 Hz. The ECG of the normal group is shown in Figure 2.

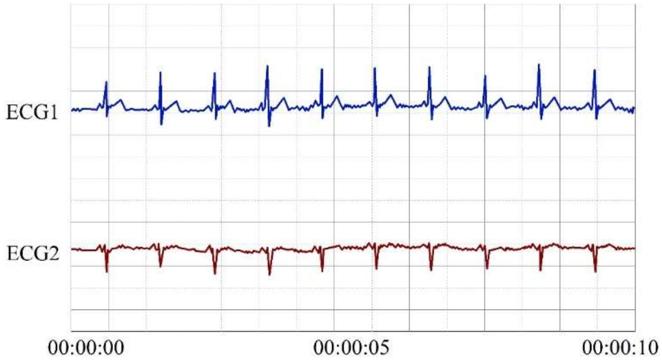


Figure 2: Electrocardiogram of the normal group

The data for the heart failure group were obtained from the BIDMC Congestive Heart Failure Database, which contains 15 long-term electrocardiogram recordings. All subjects included in the database had severe congestive heart failure (NYHA class 3-4), including 10 men (aged 22 to 71 years) and 5 women (aged 54 to 63 years). Each recording in the database lasted approximately 20 hours and included two ECG signals, each sampled at 250 times per second (i.e., a sampling rate of 250 Hz). The original analog recordings were obtained using an outpatient ECG recorder. The ECG of the heart failure group is shown in Figure 3.

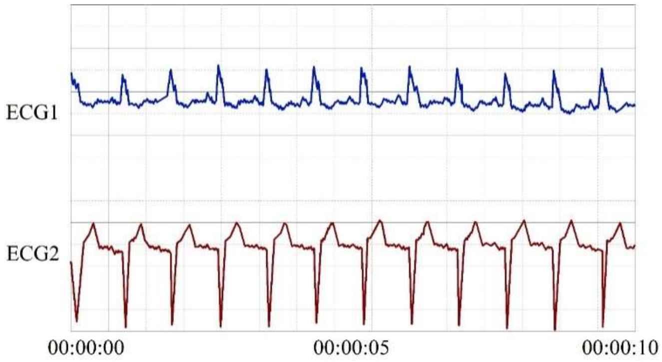
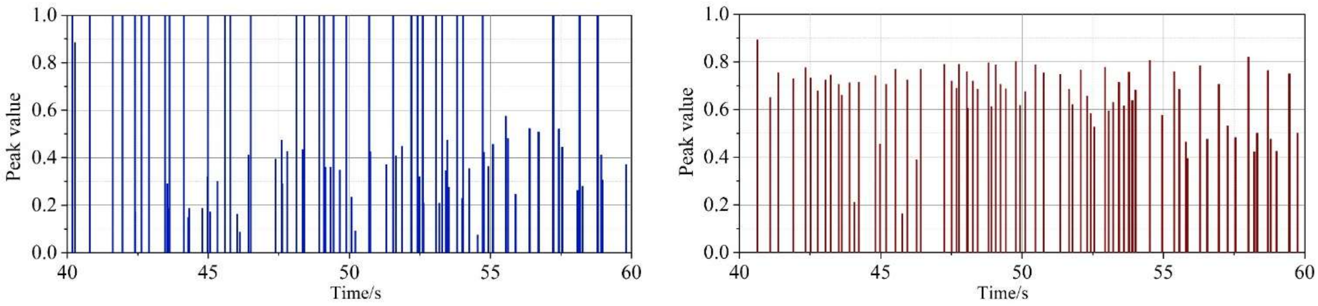


Figure 3: Electrocardiogram of the heart failure group

IV. A. 2) ECG signal preprocessing

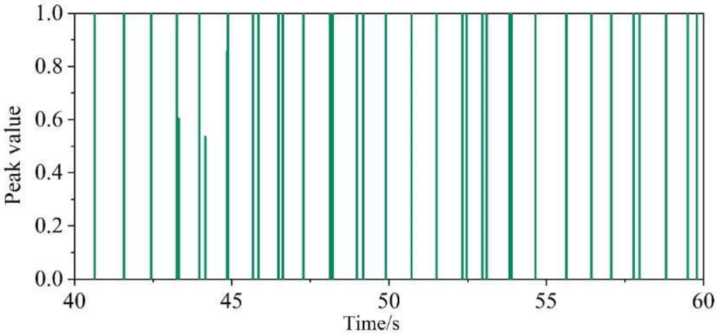
(1) R-wave detection based on multi-feature probability methods

Since HRV reflects changes in the differences between successive heartbeat cycles in the ECG signal, HRV analysis requires first calculating the ECG signal period, i.e., the RR interval sequence, from the ECG signal. Therefore, R-wave detection is necessary, followed by further calculation of the RR interval. Through steps such as preprocessing, feature extraction, probability density estimation, probability representation, and decision fusion, R-wave detection of the ECG signal is completed, laying the foundation for subsequent signal quality assessment. This method maintains good performance even under noisy conditions. Preprocessing: Since ECG signals are prone to noise during acquisition, filters must be designed for noise reduction. Here, a bandpass filter is primarily used. Since the energy of the QRS complex primarily lies between 5 Hz and 20 Hz, a bandpass filter with a bandwidth of 5 Hz to 20 Hz is designed. The filtered ECG signal is denoted as SA_ECG. After filtering, a derivative FIR filter is used to amplify the slope information of the QRS complex to distinguish the complex. Next, the signal is squared point-by-point using a nonlinear quadratic function to enhance the main peak. Finally, a smoothing filter (length 101 ms) is applied, and the signal is denoted as SF_ECG. Corresponding local maxima can be detected on SF_ECG, and the time corresponding to this maximum is recorded as t. This local maximum can be considered a potential R-wave candidate. The probability density of the R wave peak features is shown in Figure 4 (Figures a–c represent the probability density of the correlation value c, the probability density of the original signal amplitude a, and the probability density of the squared slope s, respectively).



(a) The probability density of the correlation value c

(b) The probability density of the amplitude of the original signal



(c) The probability density of the squared slope s

Figure 4: R peak eigenprobability density

Use the method to detect the R wave. The R wave detection results are shown in Figure 5. Use the method to detect the R wave. The R wave detection results are shown in Figure 5.

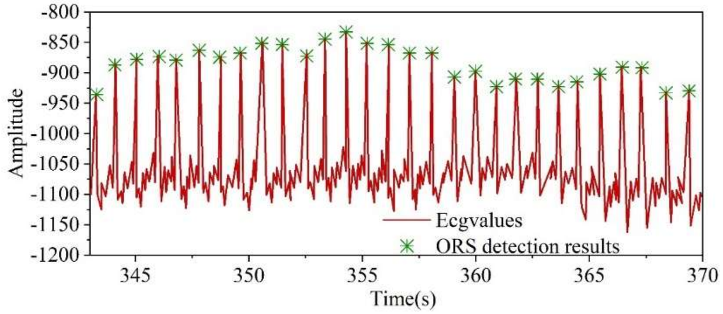


Figure 5: R-wave detection result

To better validate the performance of this method under different noise pressure conditions, noise recordings from the MIT-BIH Noise Stress Test Database (MIT-BIH Noise Stress Test Database) were added to the MIT-BIH Arrhythmia Database (MIT-BIH Arrhythmia Database) to create a baseline noise database. The R-wave detection results of the ECG signals with added noise are shown in Figure 6. It can be seen that R-wave detection remains effective even in noisy segments.

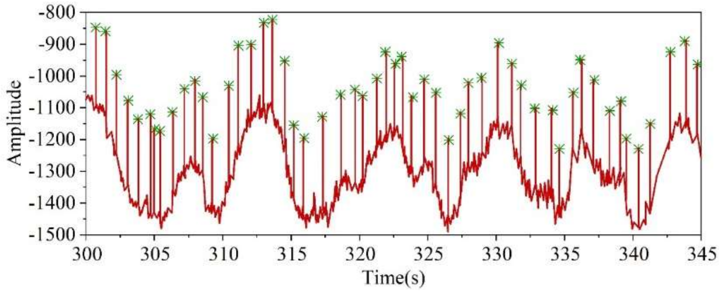


Figure 6: Detection results of R waves in ECG signals with added noise

(2) ECG signal quality assessment

Based on the above idea, the intermediate variable outputs during the detector detection process, including three features, corresponding posterior probabilities, and corresponding KLD, are used to train an SVM model with the intermediate variables and detection results. The size of the Hinge Loss corresponding to each point is calculated, and the Hinge Loss results obtained using SVM are shown in Figure 7. The average value every 30 seconds is used as the final result. As shown in the figure, there is a significant difference between the noise segment and the non-noise segment. In the noise segment, the Hinge Loss values are relatively large and scattered, while in the non-noise segment, the values are consistent and evenly distributed. According to statistics, the coefficient of variation can be used as an evaluation criterion. The coefficient of variation is a relative measure obtained by dividing the standard deviation by the mean, reflecting the variability and dispersion of the data. It effectively avoids the influence of measurement units. When the coefficient of variation exceeds 0.15, the segment can be identified as a noise segment and excluded. This feature can serve as the basis for the final signal quality assessment.

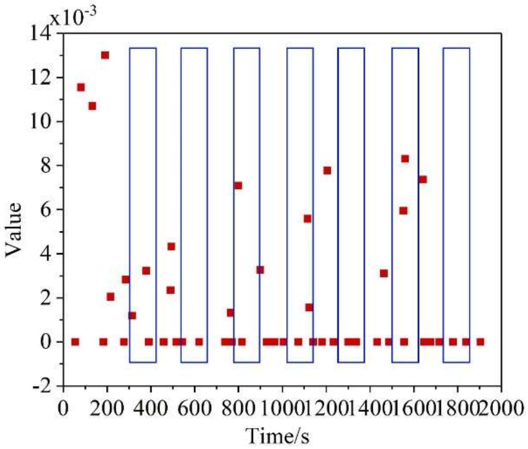


Figure 7: Use SVM for hinge loss results

IV. B. Experimental results of short-term heart sound feature analysis

IV. B. 1) Analysis results of short-term time domain characteristics of heart sounds in patients with chronic heart failure

(1) Analysis of differences in short-term time domain characteristics of heart sounds between patients with chronic heart failure and healthy individuals

The results of the calculation of short-term time domain characteristics of heart sounds in the heart failure group and the healthy group are shown in Table 1. Mean value tests were used to distinguish whether the characteristic values of the healthy group and the heart failure group had statistical differences. A P value of less than 0.05 was considered to indicate a statistical difference. Analysis of the table results shows that the mean values of D/S and S1/S2 are lower in the heart failure group (i.e., the experimental group) and higher in the healthy group (i.e., the control group), and there are statistically significant differences in heart sound short-time domain features between the heart failure group and the healthy group. The pattern of change in short-term time-domain characteristics of heart sounds indicates that, compared to healthy individuals, CHF patients exhibit significantly reduced D/S and S1/S2 values. Previous studies have demonstrated that D/S serves as a non-invasive indicator of cardiac reserve, capable of evaluating coronary artery blood supply and reflecting basic cardiac function. Therefore, it can be concluded that the onset of chronic heart failure leads to reduced cardiac reserve in CHF patients, resulting in corresponding impairment of coronary artery blood supply capacity. S1/S2 reflects the relative relationship between S1 amplitude and S2 amplitude, while S1 amplitude and S2 amplitude are respectively associated with myocardial contractility and peripheral circulatory resistance.

Table 1: The result of heart sound short-time time domain features

Group	Heart type	The number of heart sound samples.	D/S	S1/S2
Control group	Normal heart	N=300	1.836±0.205	1.679±0.192
Experimental group	CHF heart	N=700	1.491±0.195	1.177±0.233
P value			<0.01	<0.01

(2) Relationship between heart sound short-time domain characteristics and chronic heart failure staging The results of short-term time-domain characteristics of heart sounds at different stages of heart failure are shown in Table 2. It can be seen that from stage A to stage D of heart failure, as the severity of heart failure increases, the mean D/S ratio decreases progressively, and the mean S1/S2 ratio also shows a decreasing trend, indicating that the more severe the heart failure, the lower the cardiac reserve of the patient. As heart failure progresses, the D/S and S1/S2 values of heart sounds in CHF patients show a significant decreasing trend, indicating that as the severity of heart failure increases, the cardiac reserve and cardiac function of CHF patients will further decrease, and myocardial contractility will also further weaken.

Table 2: The result of heart sound short-time time domain features

Group	Heart type	The number of heart sound samples.	D/S	S1/S2
Control group	Normal heart	N=300	1.821±0.175	1.657±0.135
Experimental group	Stage A heart failure	N=150	1.643±0.142	1.217±0.17
	Stage B heart failure	N=180	1.489±0.181	1.226±0.358
	Stage C heart failure	N=210	1.39±0.201	1.154±0.287
	Stage D heart failure	N=160	1.328±0.172	1.084±0.366

IV. B. 2) Analysis results of short-term frequency domain characteristics of heart sounds in patients with chronic heart failure

(1) Analysis of differences in short-term frequency domain characteristics of heart sounds between patients with chronic heart failure and healthy individuals

This section provides a detailed description of the computational results for short-time frequency domain characteristics of heart sounds in patients with chronic heart failure and healthy individuals. The results for low-frequency energy fraction, high-frequency energy fraction, and spectral energy fraction of heart sounds in the heart failure group and healthy group are shown in Table 3. The mean low-frequency energy fraction of heart sounds in the control group (healthy group) was higher than that in the experimental group (heart failure group), while the mean high-frequency energy fraction of heart sounds in the control group was lower than that in the experimental group. It can be observed that the energy of normal heart sounds is more concentrated in the low-frequency range below 200 Hz, manifested as higher low-frequency energy scores and lower high-frequency energy scores. In contrast, the energy of signals below 200 Hz is relatively reduced in CHF heart sounds, manifested as lower low-frequency energy scores and higher high-frequency energy scores. Since the frequency components of S1 and S2 are primarily distributed below 200 Hz, and the energy of S1 is higher than that of S2.

Table 3: Calculation result of energy fraction

Group	Heart type	The number of heart sound samples.	Low-frequency energy fraction	High-frequency energy fraction	Spectral energy fraction
Control group	Normal heart	N=300	5.372±5.452	4.914±4.693	3.452±3.351
Experimental group	CHF heart	N=700	12.08±12.112	12.09±12.009	8.721±8.631
P value			<0.05	<0.05	<0.05

The results of the calculation of heart sound subband energy scores for the heart failure group and the healthy group are shown in Table 4. As can be seen, the energy of heart sounds in healthy patients is more concentrated in IMF1 and IMF2, and the subband energy score of IMF2 is the highest among all IMFs, indicating that the subband energy of normal heart sounds is primarily concentrated in IMF2. In contrast, the subband energy of CHF heart sounds is more dispersed, suggesting that the energy of normal heart sounds is more concentrated in the low-frequency range compared to CHF heart sounds. The subband energy scores of IMF1 and IMF2 are higher in the healthy group and lower in the heart failure group, indicating that the signal energy of normal heart sounds is more concentrated in the low-frequency range. Conversely, the subband energy scores of IMF3 and IMF4 are higher in the heart failure group and lower in the healthy group, indicating that during heart failure, the low-frequency energy of heart sounds decreases while the high-frequency energy increases.

Table 4: Calculation results of the energy fraction of the heart sound subband

Heart type	The number of heart sound samples.	Subband energy fraction			
		IMF1	IMF2	IMF3	IMF4
Normal heart	N=300	4.287±0.642	36.972±5.159	22.261±4.983	9.368±2.102
CHF heart	N=700	3.167±0.545	33.446±4.15	24.327±4.354	15.324±2.643
P value		<0.05	<0.05	0.07	<0.01

(2) Relationship between short-term frequency domain characteristics of heart sounds and stages of chronic heart failure

The results of calculations for low-frequency energy fraction, high-frequency energy fraction, and spectral energy fraction of heart sounds at different stages of heart failure are shown in Table 5. From stage A to stage D of heart failure, the low-frequency energy fraction of heart sounds in CHF patients shows an overall decreasing trend, while the high-frequency energy fraction shows an overall increasing trend. This indicates that as the severity of heart failure increases, the intensity of S1 weakens, myocardial contractility decreases, and changes in cardiac hemodynamic factors lead to alterations in the relative relationship between S1 and S2. As the severity of heart failure increases, the mean low-frequency energy fraction decreases progressively, while the mean high-frequency energy fraction increases progressively.

Table 5: The energy fraction of different heart failure stages

Group	Heart type	The number of heart sound samples.	low frequency energy fraction	high frequency energy fraction	spectral energy fraction
Control group	Normal heart	N=300	91.41±5.378	8.681±4.911	89.73±3.433
Experimental group	Stage A heart failure	N=150	81.93±7.638	14.59±6.866	84.82±4.334
	Stage B heart failure	N=180	76.6±8.148	21.91±7.568	77.77±5.639
	Stage C heart failure	N=210	68.96±11.114	29.55±9.877	73.46±4.257
	Stage D heart failure	N=160	63.89±11.315	35.12±13.096	65.47±7.078

The results of calculating the energy scores of heart sound subbands at each stage of heart failure are shown in Table 6. From stage A to stage D of heart failure, the energy scores of IMF1 and IMF2 subbands show a decreasing trend, while those of IMF3 and IMF4 subbands show an increasing trend. This indicates that as the severity of heart failure increases, the energy of heart sound signals tends to concentrate at higher frequencies, and the low-frequency energy of heart sound signals decreases, suggesting a reduction in the intensity of S1 and S2.

Table 6: The result of heart sound heart sound sub-band energy fraction feature

Heart type	The number of heart sound samples.	Subband energy fraction			
		IMF1	IMF2	IMF3	IMF4
Normal heart	N=300	0.639±0.645	5.171±5.161	4.98±4.983	2.112±2.105
Stage A heart failure	N=150	0.373±0.38	3.681±3.669	3.499±3.501	1.869±1.868
Stage B heart failure	N=180	0.489±0.488	4.31±4.284	4.682±4.688	2.142±2.147

Stage C heart failure	N=210	0.516±0.505	4.782±4.791	3.982±3.97	3.521±3.509
Stage D heart failure	N=160	0.608±0.609	5.598±5.576	6.432±6.417	3.784±3.779

IV. B. 3) Analysis results of short-term nonlinear characteristics of heart sounds in patients with chronic heart failure

(1) Analysis of differences in short-term nonlinear characteristics of heart sounds between patients with chronic heart failure and healthy individuals

This section provides a detailed description of the calculation results for short-term nonlinear characteristics of heart sounds between patients with chronic heart failure and healthy individuals. The calculation results for short-term nonlinear characteristics of heart sounds in the heart failure group and the healthy group are shown in Table 7. The healthy group exhibited higher mean values for heartbeat sample entropy and multiple fractal spectrum width, while the heart failure group exhibited lower mean values for heartbeat sample entropy and multiple fractal spectrum width. Additionally, statistically significant differences in short-term nonlinear characteristics were observed between the heart failure group and the healthy group. The heart sound sample entropy and multiple fractal spectrum width in the heart failure group showed an overall decreasing trend compared to the healthy group, indicating that during the onset of chronic heart failure, the chaos, complexity, and disorder of heart sound signals decrease. Since heart sounds directly reflect cardiac mechanical activity, the intrinsic nonlinear characteristics of heart sound signals can be used to describe the nonlinear properties of the cardiac mechanical dynamic system. Therefore, it can be concluded that the chaos, complexity, and disorder of cardiac mechanical activity decrease during the chronic phase.

Table 7: The result of heart sound short-time nolinear features

group	Heart type	The number of heart sound samples.	Sample entropy	Multifractal spectral width
Control group	Normal heart	N=300	0.015±0.002	1.734±0.275
Experimental group	CHF heart	N=700	0.007±0.003	1.264±0.322
P value			<0.05	<0.05

(2) Relationship between short-term nonlinear characteristics of heart sounds and chronic heart failure staging

This section investigates the relationship between short-term nonlinear characteristics of heart sounds and chronic heart failure staging by analyzing the differences in these characteristics among patients at different stages of heart failure. The results of short-term nonlinear characteristic calculations for heart sounds at different stages of heart failure are shown in Table 8. It can be observed that, from stage A to stage D of heart failure, as the severity of heart failure increases, the mean values of sample entropy and multifractal spectrum width decrease progressively. As the course of heart failure progresses, the significant decrease in sample entropy and multiple fractal spectrum width of heart sounds in CHF patients indicates that as the severity of heart failure increases, the chaos, complexity, and disorder of cardiac mechanical activity decrease.

Table 8: The result of heart sound short-time nolinear features

Heart type	The number of heart sound samples.	Sample entropy	Multifractal spectral width
Normal heart	N=300	0.012±0.006	1.73±0.277
Stage A heart failure	N=150	0.016±0.001	1.625±0.182
Stage B heart failure	N=180	0.004±0.001	1.227±0.184
Stage C heart failure	N=210	0.007±0.005	1.146±0.205
Stage D heart failure	N=160	0.002±0.002	1.002±0.24

V. Heart failure prediction based on multi-scale one-dimensional convolutional recurrent neural networks

V. A. Convolutional Neural Networks

V. A. 1) Basic Theory of Convolutional Neural Networks

CNNs are composed of filters, kernels, or neurons with learnable weights, parameters, and biases. The main differences between CNNs and other neural networks lie in two aspects: local connections and weight sharing. These features result in fewer internal weights within the CNN, reducing model complexity and lowering the risk of overfitting. A typical CNN architecture consists of three components: convolutional layers, pooling layers, and fully connected layers.

(1) Convolution Layer

The convolution layer is the core module of the network, performing most of the computational work, with the primary purpose of feature extraction. Convolution operations learn features from the input image while preserving spatial relationships between pixels. The convolution layer consists of multiple convolution kernels, which perform convolution on the input image using a set of learnable neurons, generating different feature maps in the output image. These feature maps are then fed as input data to the next convolution layer. Specifically, each neuron in the output feature map is locally connected to its input receptive field. The input is convolved with a learned kernel, a bias term is added to the convolution result, and finally a nonlinear activation function is applied to obtain a new feature map, as shown in Formula:

$$z_{i,j,k}^{l} = f\left(w_{k}^{lT} x_{i,j}^{l} + b_{k}^{l}\right) \tag{25}$$

In the equation, $z_{i,j,k}^{l}$ is the feature value at position (i,j) of the k th feature map in the l th layer, w_{k}^{l} is the weight of the k th feature map in the l th layer, and b_{k}^{l} is the bias of the k th feature map in the l th layer. T is the size of the convolution kernel, and $f(\cdot)$ is a nonlinear activation function.

Activation functions introduce nonlinear factors into neural networks, enhancing the expressive power of linear models and enabling neural networks to be applied to more nonlinear models. Commonly used activation functions include $sigmoid$, $\tanh$, and $\mathrm{Re}LU$.

1) $sigmoid$ function

The mathematical expression for the $sigmoid$ function is as follows:

$$f(x) = \frac{1}{1+e^{-x}} \tag{26}$$

2) $\tanh$ function

The mathematical expression for the $\tanh$ function is as follows:

$$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{27}$$

The $\tanh$ function solves the problem of the $sigmoid$ function outputting a non-zero mean, but from its function graph, we can see that when the input tends to positive infinity or negative infinity, the output is smooth, the gradient is very small, and the gradient vanishes. Similarly, the $\tanh$ function also involves power operations, which are computationally expensive and slow down training.

3) $\mathrm{Re}LU$ function

The mathematical expression for the $\mathrm{Re}LU$ function is as follows:

$$f(x) = \max(0,x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{28}$$

The $\mathrm{Re}LU$ function has only linear relationships, is simple to compute, and is much faster than the $sigmoid$ and $\tanh$ functions, making it one of the most widely used activation functions today. However, from the function's graph, it still has some drawbacks: on one hand, when $x < 0$, the output is 0, and the $\mathrm{Re}LU$ function is completely deactivated, meaning that certain neurons will never be activated, and the corresponding parameters will not be updated. On the other hand, the output of the $\mathrm{Re}LU$ function is not zero-mean. To address the shortcomings of $\mathrm{Re}LU$, some improved versions have been proposed, such as ELU and $P\mathrm{Re}LU$. Both have a certain output when the input is negative, solving the “death” phenomenon of $\mathrm{Re}LU$, but they both increase the computational load due to the introduction of hyperparameters.

(2) Pooling layer

The pooling layer, also known as the downsampling layer, reduces the dimension of the feature map and reduces the number of computational parameters while retaining the original features. It is typically placed between convolutional layers. The pooling layer has good generalization performance, converges quickly, and is less prone to translation and distortion. The pooling computation process is as follows:

$$y_{i,j,k}^{l} = pooling(a_{m,n,k}^{l}), \forall(m,n) \in R_{ij} \tag{29}$$

In the equation, $pooling(\cdot)$ denotes the pooling function, $a_{m,n,k}^{l}$ represents the feature map after convolution, and R_{ij} denotes the local neighborhood at position (i, j).

Common pooling methods include max pooling, average pooling, and random pooling. The pooling operation diagram is shown in Figure 8, where the input feature map size is 4×4, the convolution kernel size is 2×2, and the stride is 2. When using the pooling operation, the feature map is first divided into 4 different colored 2×2 regions. Max pooling takes the maximum value of the feature points within the 2×2 region as the pooled value, and this method is sensitive to texture features. Mean pooling takes the average value of the feature points within the 2×2 region as the pooled value, and this method is beneficial for retaining background information. Random pooling assigns probability values to feature points within the 2×2 region based on their magnitude, then randomly selects values from the feature map based on the magnitude of these probability values. This method assigns a probability value to each neuron in the feature map, ensuring that every neuron has the potential to be utilized while also avoiding overfitting.

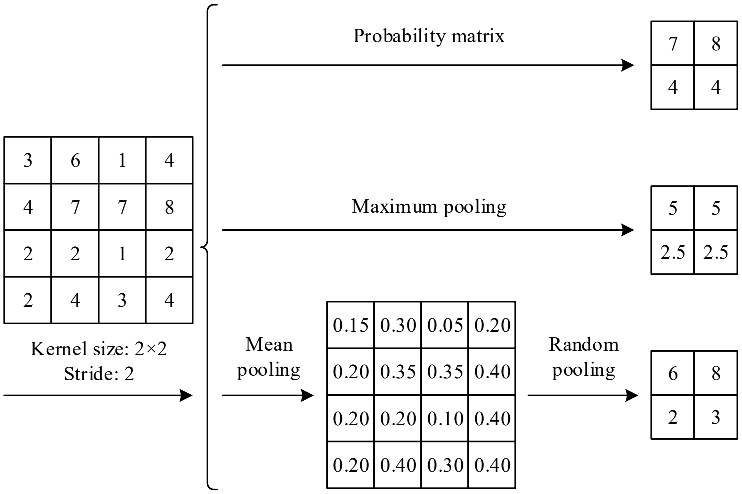


Figure 8: The illustration of pooling operations

(3) Fully Connected Layer

In a CNN structure diagram, after multiple convolution layers and pooling layers, there is usually a fully connected layer that integrates the output features of the convolution layers and pooling layers to classify the input into various categories. Each neuron in the fully connected layer is connected to all neurons in the previous layer. The last layer is connected to the output layer and passes the features to the output layer, where a softmax classifier is used to complete the target classification.

The calculation for the fully connected layer is as follows:

$$\begin{aligned} a_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + b_1 \\ a_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3 + b_2 \\ a_3 &= w_{31}x_1 + w_{32}x_2 + w_{33}x_3 + b_3 \end{aligned} \tag{30}$$

In this case, x_1, x_2, and x_3 are the inputs to the fully connected layer, a_1, a_2, and a_3 are the outputs, and w and b represent the weights and biases, respectively.

V. A. 2) Convolutional Neural Network Training Strategy

The training method for CNNs is the backpropagation algorithm, which consists of two stages: forward propagation and backward propagation. During forward propagation, training data passes through the input layer and hidden layers before finally producing an output result at the output layer. There is a certain error between the network's output result and the true value. This error is propagated backward from the output layer until it reaches the input layer. During backward propagation, the values of various parameters are continuously adjusted to make the output result closer to the true value. Parameter optimization in neural networks involves methods such as gradient descent, momentum gradient descent, and adaptive learning rate to make the network's output values closer to the true values.

V. B. Multi-scale one-dimensional convolutional neural networks

V. B. 1) One-dimensional convolutional neural networks

Heart sound signals are one-dimensional signals. Using traditional CNNs to classify and identify them requires dimensional transformation, which increases the complexity of the algorithm and may result in the loss of useful information during the conversion process. Therefore, in order to make CNNs suitable for processing one-dimensional heart sound data, they need to be converted into a one-dimensional form.

In traditional CNNs, assume that $X \in R^{H\times W\times D}$ represents the input image or feature map, where H , W , and D represent the height, width, and number of feature maps, respectively. The input generated after the convolution operation is as follows:

$$s = \sum_{i,j,m} x_{i,j,m} w_{k,i,j,m} \tag{31}$$

In the equation, $i = 1,2,...,h$ · $j = 1,2,...,w$ · $m = 1,2,...,D$ · $w_k \in R^{h\times w\times D}$ denotes the size of the convolution kernel, and $x \in R^{h\times w\times D}$ denotes a spatial block belonging to X , with the same size as the convolution kernel w_k . The corresponding output feature map f_k can be expressed as:

$$f_k = X * w_k \in R^{H\times W\times D} \tag{32}$$

Here, * denotes convolution operation.

When the input of CNN is a one-dimensional signal, the input $X' \in R^{H\times 1\times D}$ represents D one-dimensional feature vectors, each of which has a size of $H \times 1$. Correspondingly, the convolution kernel is represented as $w' \in R^{h\times 1\times D\times D'}$, with the feature map width w set to 1 . Specifically, if the number of input channels D in the first layer is 1 , then the convolution window must also be adjusted to $h \times 1$. Thus, the output feature vector f_k' is represented as:

$$f_k' = X' * w_k' \in R^{H\times 1\times 1}, k = 1,2,...,D' \tag{33}$$

V. B. 2) Multi-scale convolution

Multi-scale convolution utilizes convolution kernels of different sizes to achieve complementary information at different scales. Multi-scale blocks can generate receptive fields of different sizes to capture information at different scales. The smaller the scale of the convolutional kernel, the more detailed the information captured. Conversely, the larger the scale, the more abstract the features captured. Combining convolutional kernels of different scales allows both detailed and abstract information to be retained simultaneously, which also enhances the feature representation capability of CNNs and improves classification performance.

V. C. Network Structure Design

This section combines MSCNN with MGU to build a heart failure classification model, using CNN for feature extraction and the MGU network to learn the temporal features of the sequence. The model proposed in this chapter is named Multi-Scale-CNN-MGU (MSCNN-MGU), and the structure of the MSCNN-MGU model is shown in Figure 9.

VI. Model performance testing experiment

This section compares and analyzes the performance of the model under different experimental conditions. To ensure the reliability of the experimental results, all models in the experiments were trained using the same parameters and methods, as described below. During model training, Cross Entropy was used as the loss function. Considering the GPU's 24GB of VRAM, the Batch Size was set to 128. Using the largest possible Batch Size helps avoid model divergence caused by bad data, provides some regularization effects, suppresses overfitting, and improves its generalization ability. During model initialization, EfficientNet uses pre-trained weights from ImageNet_1K. The convolutional layers and fully connected layers involved in feature fusion are initialized with random parameters. During training, six threads are enabled to read ECG data in parallel and perform preprocessing. For each training session, the model iterates 500 epochs and saves the optimal weight values. Considering the scale of the dataset, the optimizer selected Adam. The initial learning rate is set to 1×10^{-4}, and as the training process progresses, the *RRRRmmRRAARR LLRRRRttffll ffttRRffRR* algorithm for adaptive adjustment. After detecting that the loss no longer decreases or the evaluation metric values no longer increase over 10 epochs, the learning rate is automatically reduced by 0.8, maximizing the alignment between the learning rate and the model iteration process, enabling it to fully learn the knowledge in the dataset. The decrease in the learning rate and loss convergence during a training process are shown in Figure 10 (Figure a shows the learning rate curve, and Figure b shows the loss value changes). The horizontal axis of both figures represents the epoch, while the vertical axes represent the learning rate and loss value, respectively, demonstrating that the model converges gradually during training.

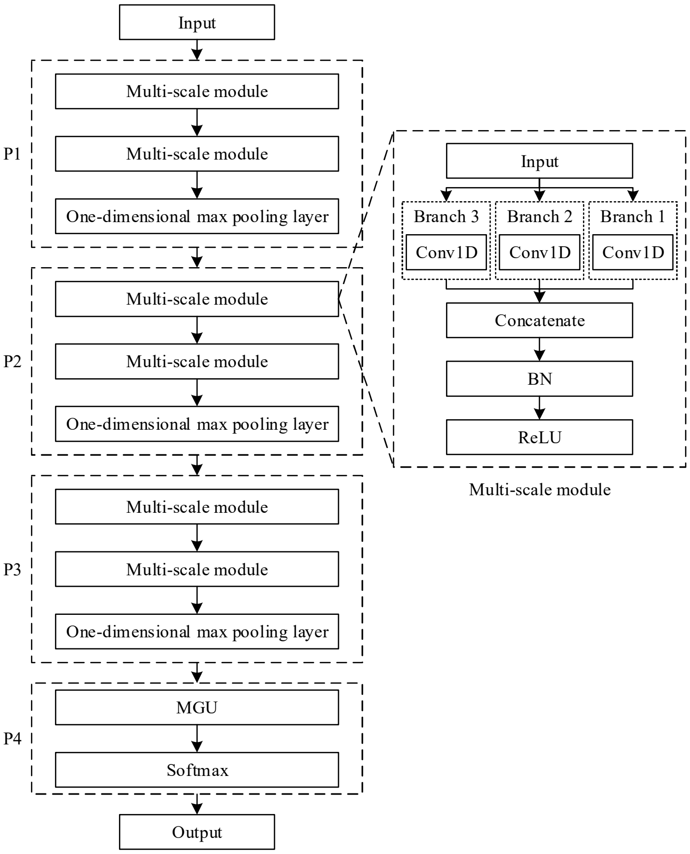
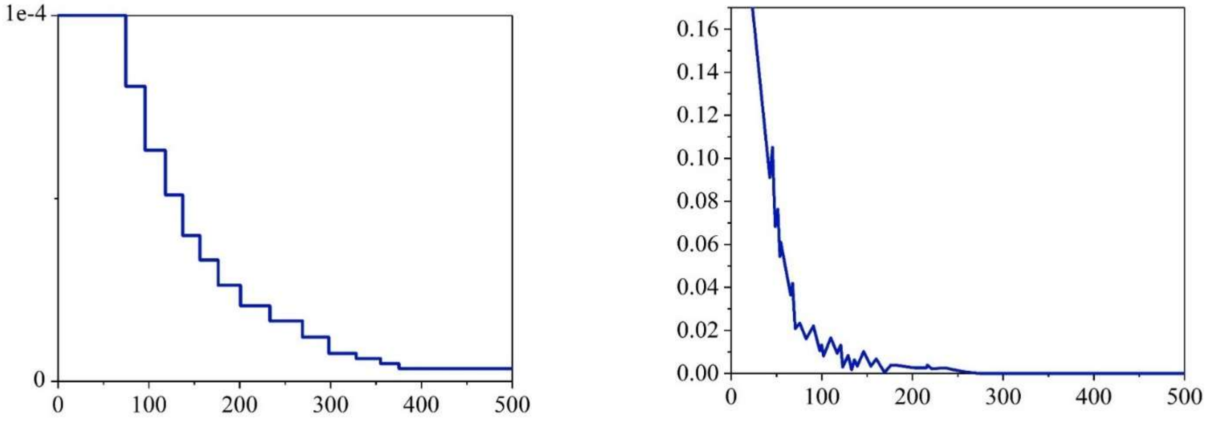


Figure 9: The structure of the MSCNN-MGU model



(a) Learning rate curve

(b) Change in loss value

Figure 10: The learning rate decreases and the loss converges

VI. A. Handling unbalanced datasets

The purpose of model training is to minimize the loss function. When the number of samples in a particular category is extremely large, the value of the loss function is predominantly influenced by the category with the larger sample size, resulting in the model tending to classify toward the category with the larger sample size. The dataset used in this paper is a classic imbalanced dataset. The kernel density of the LVEF values was calculated based on 56,817 data points. The kernel density estimation plot of the dataset shows that the data is primarily distributed in the range of 0.5 to 0.8, with a severe imbalance between positive and negative samples and a significant long-tail problem.

Combining oversampling and undersampling methods, the model performs random undersampling on negative samples in the dataset during training and oversampling on positive samples. Positive samples in the dataset are sampled multiple times to ensure that the information in positive sample data is maximized. At the same time, negative samples four times the number of positive samples are randomly selected for training to avoid weakening the model's focus on positive sample features due to an excessive proportion of negative samples. To validate the method, the model was trained before and after applying the imbalanced dataset processing method. The F1 values of the positive samples after processing the imbalanced data are shown in Figure 11. The red line represents the training results before processing, with an F1 value hovering around 50%. The blue line represents the training results after processing, with the F1 score converging above 80%. It is evident that training the imbalanced dataset using oversampling and random undersampling methods effectively improved the F1 metric by nearly 30%. When training the original dataset in its entirety, the long tail prevents the model from effectively learning feature information from the weak positive samples, and the gradients generated by the loss function fail to prompt the model to effectively fit the ECG features of heart failure patients. The processed dataset is well “digested” by the model during training, and as the number of iterations increases, the model gradually acquires the knowledge to distinguish whether a patient's ECG indicates heart failure.

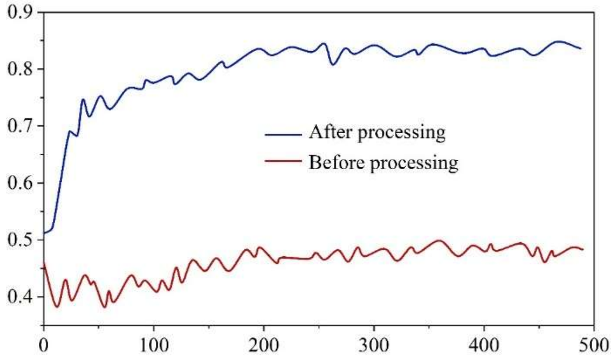


Figure 11: Comparison of the F1 values of the positive samples

VI. B. Data Augmentation and Its Effectiveness Experiment

In deep learning, data augmentation is a commonly used technique in the field of computer vision. By applying methods such as flipping, cropping, and scaling images in a dataset, it is possible to expand the training dataset and improve the model's generalization ability without substantially increasing the amount of data, thereby enabling limited data to yield value equivalent to that of a larger dataset. Data augmentation can also be applied to ECG signal data. In the dataset used for the experiment, due to the difficulty in collecting data, the proportion of positive samples is significantly smaller than that of negative samples, resulting in a limited number of positive samples. This makes the model prone to bias and overfitting during the learning process. Common methods for data augmentation on signal data include random perturbation, signal transformation, signal synthesis, and signal interpolation. Considering the periodic characteristics of ECG signals, this paper performed random segmentation of multi-lead ECG data with a window length of 4 seconds. The starting position of each segmentation follows a normal distribution, and Gaussian background noise is introduced into the fused spectrum to fill gaps in certain segments, while random drift is also added. These operations expand the dataset and enhance the algorithm's robustness to different types of noise and deformations. The impact of data augmentation on model performance is shown in Figure 12. The horizontal axis represents the number of model iterations, and the vertical axis represents the F1 score. The red curve shows the training process of the model without data augmentation, while the blue curve shows the training process of the model with data augmentation. The model using the unaugmented dataset rises rapidly during training but enters overfitting early on, with the F1 score eventually converging around 84%. The model using the augmented dataset experiences persistent perturbations during the learning process, which although slowing down the model's convergence speed in the early stages of training, also reduces the extent of overfitting and increases the convergence height in the later stages, with the F1 value during training exceeding 85%. It is evident that the application of data augmentation provides at least a 1% performance improvement, and training with augmented data has a certain overfitting suppression effect. Among the data augmentation methods, random cutting and random shifting alter the start and end positions of the ECG signal, increasing the model's perspective on the signal. This ensures that even the same patient data segment does not remain identical across multiple iterative learning processes. Data augmentation enhances the model's learning effectiveness, preventing it from simply memorizing the characteristics of a specific patient's data during iteration while pursuing loss minimization. Instead, it forces the model to learn the commonalities across various patient data types.

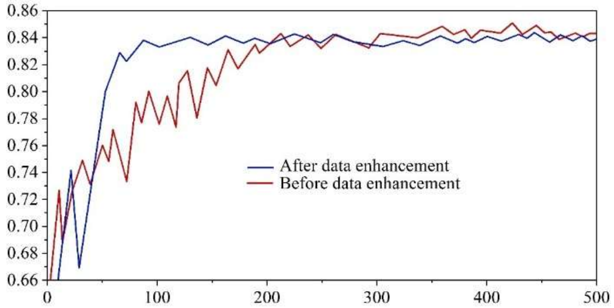


Figure 12: The impact of data augmentation on model performance

VI. C. Performance Comparison Between Different Models

This section compares the performance differences of several neural network models that have demonstrated strong performance in the field of computer vision on the dataset, including AlexNet, VGG, ResNet, DenseNet, and Vit. Performance Comparison of Different Neural Network Models As shown in Table 9, the results indicate that for the classification task on the dataset in this chapter, EfficientNet-B2 achieves the following performance metrics: Acc=94.13, Pre=92.38, Rec=81.93, F1=84.5, AUC=0.945, with an inference speed of 917ms when using a BatchSize of 128. Thus, while EfficientNet-B2 has similar model performance to Densenet-161, it demonstrates higher balanced performance while also ensuring faster inference speed.

Table 9: Performance comparison of different neural network models

Model	Acc (%)	Pre (%)	Rec (%)	F1 (%)	AUC	Time (ms)
AlexNet	86.49	75.11	82.77	78.73	0.903	155
VGG16	87.83	79.61	77.8	82.71	0.942	602
VGG19	92.06	82.11	85.56	84.87	0.942	741
Res-18	88.39	78.78	79.34	81.17	0.911	407
Res-34	93.45	87.76	85.94	87.55	0.949	788
Res-50	88.15	85.59	84.93	81.66	0.952	1241
Res-101	90.92	86.56	82.54	87.04	0.937	2065
Res-152	93.36	86.03	83.14	85.06	0.942	2372
Densenet121	95.72	85.2	82.7	83.71	0.942	827
Densenet161	93.42	92.71	84.12	88.88	0.969	1640
Densenet169	96.03	79.85	84.31	84.17	0.955	1021
Densenet201	92.11	80.85	86.56	84.26	0.952	1278
EfficientNet-b1	93.43	83.28	83.9	85.26	0.95	646
EfficientNet-b2	94.13	92.38	81.93	84.5	0.945	917
EfficientNet-b3	93.17	83.29	85.11	82.36	0.952	1177
EfficientNet-b4	93.92	85.15	83.65	82.03	0.947	1456
Vit-16	88.4	81.47	81.4	79.48	0.765	1806
Vit-32	90.04	84.09	76.2	80.31	0.87	555

VI. D. Neural Network Visualization and Feature Pruning

The performance of the model detection under different numbers of leads is shown in Table 10. Single-lead, three-lead, six-lead, and twelve-lead ECGs were selected as inputs to evaluate the performance of the algorithm model in this chapter. As the number of leads decreases, the model performance decreases slightly, but not significantly, indicating that the model can adapt well to input data with different numbers of leads.

Table 10: The performance of model checking under different lead numbers

Number of leads	Acc (%)	Pre (%)	Rec (%)	F1 (%)	AUC
1	89.42	83.01	79.43	76.19	0.873
3	89.47	83.86	82.75	83.71	0.906
6	94.9	86.12	79.95	84.38	0.913
12	93.17	97.16	83.18	83.25	0.976

VII. Conclusion

This paper investigates algorithms for using electrocardiogram (ECG) signals in the diagnosis of heart failure and designs a heart failure prediction model based on a multi-scale one-dimensional convolutional recurrent neural network (MSCNN-MGU). The main findings of this study are as follows:

This study investigates the application of the MSCNN-MGU heart failure prediction model in detecting left ventricular dysfunction (LVD) patients using 12-lead ECG, confirming the feasibility of analyzing ECG using this algorithmic model to assess cardiac function. For example, the performance of EfficientNet-B2 is as follows: Accuracy (Acc) = 94.13%, Precision (Pre) = 92.38%, Recall (Rec) = 81.93%, F1 score = 84.5%, AUC = 0.945, with an inference speed of 917 ms when using a BatchSize of 128, indicating that the model has a relatively fast inference speed.

This paper investigates methods for processing and analyzing physiological signal data related to heart failure patients and designs a heart failure classification model based on a multi-scale one-dimensional convolutional recurrent neural network, which has potential value for the diagnosis of heart failure and the prediction of cardiovascular diseases.

Funding

This work was supported by Science and Technology Innovation team of Heilongjiang University of Traditional Chinese medicine (2019TD01).

References

[1] Dar, T., Radfar, A., Abohashem, S., Pitman, R. K., Tawakol, A., & Osborne, M. T. (2019). Psychosocial stress and cardiovascular disease. Current treatment options in cardiovascular medicine, 21, 1-17.

[2] Chauvet-Gelinier, J. C., & Bonin, B. (2017). Stress, anxiety and depression in heart disease patients: A major challenge for cardiac rehabilitation. Annals of physical and rehabilitation medicine, 60(1), 6-12.

[3] Xanthopoulos, A., Starling, R. C., Kitai, T., & Triposkiadis, F. (2019). Heart failure and liver disease: cardiohepatic interactions. JACC: Heart Failure, 7(2), 87-97.

[4] Kalter-Leibovici, O., Freimark, D., Freedman, L. S., Kaufman, G., Ziv, A., Murad, H., ... & Silber, H. (2017). Disease management in the treatment of patients with chronic heart failure who have universal access to health care: a randomized controlled trial. BMC medicine, 15, 1-13.

[5] Menachem, J. N., Schlendorf, K. H., Mazurek, J. A., Bichell, D. P., Brinkley, D. M., Frischhertz, B. P., ... & Lindenfeld, J. (2020). Advanced heart failure in adults with congenital heart disease. Heart failure, 8(2), 87-99.

[6] Schwinger, R. H. (2021). Pathophysiology of heart failure. Cardiovascular diagnosis and therapy, 11(1), 263.

[7] Meijers, W. C., & de Boer, R. A. (2019). Common risk factors for heart failure and cancer. Cardiovascular research, 115(5), 844-853.

[8] Doehner, W., Ural, D., Haeusler, K. G., Čelutkienė, J., Bestetti, R., Cavusoglu, Y., ... & Ruschitzka, F. (2018). Heart and brain interaction in patients with heart failure: overview and proposal for a taxonomy. A position paper from the Study Group on Heart and Brain Interaction of the Heart Failure Association. European journal of heart failure, 20(2), 199-215.

[9] Nordfonn, O. K., Morken, I. M., Bru, L. E., & Husebø, A. M. L. (2019). Patients' experience with heart failure treatment and self-care—a qualitative study exploring the burden of treatment. Journal of clinical nursing, 28(9-10), 1782-1793.

[10] Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. International Journal of Advanced Computer Science and Applications, 10(6).

[11] Elgendy, I. Y., Mahtta, D., & Pepine, C. J. (2019). Medical therapy for heart failure caused by ischemic heart disease. Circulation research, 124(11), 1520-1535.

[12] Reddy, Y. N., Carter, R. E., Obokata, M., Redfield, M. M., & Borlaug, B. A. (2018). A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. Circulation, 138(9), 861-870.

[13] Castiglione, V., Aimo, A., Vergaro, G., Saccaro, L., Passino, C., & Emdin, M. (2022). Biomarkers for the diagnosis and management of heart failure. Heart failure reviews, 1-19.

[14] Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K., & Fotiadis, D. I. (2017). Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. Computational and structural biotechnology journal, 15, 26-47.

[15] Abe, H., Semba, H., & Takeda, N. (2017). The roles of hypoxia signaling in the pathogenesis of cardiovascular diseases. Journal of atherosclerosis and thrombosis, 24(9), 884-894.

[16] Jahmunah, V., Ng, E. Y. K., San, T. R., & Acharya, U. R. (2021). Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. Computers in biology and medicine, 134, 104457.

[17] Simegn, G. L., Gebeyehu, W. B., & Degu, M. Z. (2022). Computer-aided decision support system for diagnosis of heart diseases. Research Reports in Clinical Cardiology, 39-54.

[18] Hagiwara, Y., Fujita, H., Oh, S. L., Tan, J. H., San Tan, R., Ciaccio, E. J., & Acharya, U. R. (2018). Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review. Information Sciences, 467, 99-114.

[19] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Computer methods and programs in biomedicine, 141, 19-26.

[20] Loh, B. C., & Then, P. H. (2017). Deep learning for cardiac computer-aided diagnosis: benefits, issues & solutions. Mhealth, 3, 45.

[21] Yaseen, Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. Applied Sciences, 8(12), 2344.

[22] Thiyagaraja, S. R., Dantu, R., Shrestha, P. L., Chitnis, A., Thompson, M. A., Anumandla, P. T., ... & Dantu, S. (2018). A novel heart-mobile interface for detection and classification of heart sounds. Biomedical Signal Processing and Control, 45, 313-324.

[23] Gjoreski, M., Gradišek, A., Budna, B., Gams, M., & Poglajen, G. (2020). Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. Ieee Access, 8, 20313-20324.

[24] Abdollahpur, M., Ghaffari, A., Ghiasi, S., & Mollakazemi, M. J. (2017). Detection of pathological heart sounds. Physiological measurement, 38(8), 1616.

[25] Gjoreski, M., Simjanoska, M., Gradišek, A., Peterlin, A., Gams, M., & Poglajen, G. (2017, August). Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers. In 2017 international conference on intelligent environments (IE) (pp. 14-19). IEEE.

[26] Jahmunah, V., Oh, S. L., Wei, J. K. E., Ciaccio, E. J., Chua, K., San, T. R., & Acharya, U. R. (2019). Computer-aided diagnosis of congestive heart failure using ECG signals–A review. Physica Medica, 62, 95-104.

[27] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., & Tan, R. S. (2019). Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Applied Intelligence, 49, 16-27.

[28] Tripathy, R. K., Paternina, M. R., Arrieta, J. G., Zamora-Méndez, A., & Naik, G. R. (2019). Automated detection of congestive heart failure from electrocardiogram signal using Stockwell transform and hybrid classification scheme. Computer methods and programs in biomedicine, 173, 53-65.

[29] Geweid, G. G., & Abdallah, M. A. (2019). A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique. IEEE Access, 7, 149595-149611.

[30] Nogueira, M. A., Calcagno, S., Campbell, N., Zaman, A., Koulaouzidis, G., Jalil, A., ... & Kecskes, I. (2023). Detecting heart failure using novel bio-signals and a knowledge enhanced neural network. Computers in Biology and Medicine, 154, 106547.

[31] Samira Moghani,Hossein Marvi & Zeynab Mohammadpoory. (2025). Enhanced heart sound analysis through hierarchical spectral basis vector extraction using deep orthogonal non-negative matrix factorization. The Journal of Supercomputing,81(8),899-899.

[32] Wang Juliang,Zang Junbin,Yao Shuai,Zhang Zhidong & Xue Chenyang. (2024). Multiclassification for heart sound signals under multiple networks and multi-view feature. Measurement,225,114022-.