ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

Multi-feature Fusion Recognition Method of Heart Sound Signal Based on GADF and CNN-ViT

Xiong Weihua*, Jiang Yufei, Cao Lixian

School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, 132000, China
w59ulf@163.com
**Corresponding author*

Abstract: *As the most direct means to diagnose cardiovascular diseases, heart sound classification is attracting the attention of researchers all over the world. Auscultation as a traditional method, its effectiveness largely depends on the physician's clinical experience. Therefore, the development of heart sound classification and recognition in the direction of intelligence has become the mainstream trend. At present, most of the researches mainly focus on the hierarchical feature extraction of signals, which will cause the problem of insufficient feature extraction and affect the accuracy and stability of heart sound signal classification. In order to extract more comprehensive features and improve the recognition accuracy of heart sound signals, this paper constructs a multi-feature fusion recognition network. Firstly, the heart sound signal is preprocessed and converted into GADF image data. Then, the preprocessed one-dimensional signal and corresponding GADF image are input into CNN-ViT1 and CNN-ViT2 channels respectively for feature extraction. The combination of SimAM module and LeakyReLU activation function avoids the problem of "dead neurons" while enhancing the learning ability of nonlinear features and improving the ability to identify key features. Finally, the features extracted from the two channels are spliced in the channel dimension and input into the fully connected layer for classification recognition. The experimental results show that the recognition accuracy of this method is 98.81%, the sensitivity is 98.41%, the specificity is 97.86%, the accuracy is 98.84%, and the F1 score is 97.92%, which provides reliable technical support for the classification and recognition of heat sound signals. The dataset can be accessed at "https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features-/blob/master/README.md".*

Keywords: *Heart Sound Signal; Feature Fusion; Gram Angular Difference Field (GADF); Convolutional Neural Networks (CNN); Vision Transformer (ViT)*

1. Introduction

Heart sound (HS) signals are physiological signals that result from myocardial movement and the opening and closing of heart valves. They can provide useful diagnostic information about the heart and have a positive impact on the early diagnosis of cardiovascular diseases. Currently, doctors can obtain heart sound signals with the help of digital stethoscopes and then identify them through modern methods such as cutting-edge digital signal processing and machine learning techniques[1-4]. The heart sound cycle of a normal adult mainly includes the first heart sound (S1), systole (Sys), the second heart sound (S2), and diastole (Dia), as shown in Figure 1.

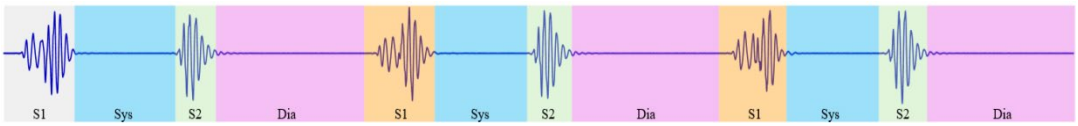


Figure 1. Schematic diagram of normal heart sound signals.

The normal contraction and relaxation of the heart are the basis of the human body's blood circulation. When there are abnormalities in the heart, they will be manifested in the heart sound signals. Currently, the prevalence and mortality rates of heart valve diseases (HDVs), including mitral valve prolapse (MVP), mitral regurgitation (MR), aortic stenosis (AS), mitral stenosis (MS), etc., are continuously increasing [5], and they have become a major threat to global human health [6-7].

In recent years, scholars worldwide have advanced heart sound signal recognition using deep

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

learning. Milani et al. [8] proposed an LDA/ANN classification model, but its generalization was limited due to LDA's inter-class difference maximization-based dimensionality reduction. Mei et al. [9] developed a method combining quality assessment and waveform scattering transformation, yet underutilized time-frequency information. Chen et al. [10] created a spectrogram-based robust deep learning framework, demonstrating high accuracy and robustness in classification tasks. Wang et al. [11] innovated the LCACNN model by fusing Mel frequency spectrum coefficients and envelope features, achieving 91.78% and 94.79% accuracies on PhysioNet and HS databases respectively. Zhang H et al. [12] proposed MDFNet with multi-dimensional feature extraction and decision fusion modules, excelling in binary and five-class classification. Zhang X et al. [13] combined MFCCs and bispectral features but neglected intrinsic time-frequency information, compromising performance on complex signals. Current trends indicate that integrating multi-feature fusion with deep models effectively improves classification accuracy, with full utilization of time-frequency information key to solving complex signal recognition challenges.

Based on the above analysis, in order to further improve the recognition accuracy of heart sound signals, this paper constructs a multi-feature fusion network combining CNN and ViT for heart sound signal recognition. Firstly, the one-dimensional signal is preprocessed and then transformed into a GADF image. Secondly, the preprocessed one-dimensional signal and the corresponding GADF image are used as inputs and fed into the CNN-ViT1 and CNN-ViT2 channels of the multi-feature fusion network respectively for feature extraction. Among them, the SimAM module and the LeakyReLU activation function are added to each channel. By reallocating the weights of different features of the heart sound signal and enhancing the nonlinear learning ability of the model, the network's ability to recognize key signal features is enhanced. Finally, the features extracted from the two channels are concatenated according to the channel dimension and input into the fully connected layer for classification and recognition. At the end of this paper, through multiple comparative experiments, it is further verified that the method proposed in this paper has the best overall performance, which improves the recognition accuracy of heart sound signals.

2. Relation Work

2.1 Gram Angular Difference Field (GADF)

Time series is a one-dimensional signal. In the Cartesian coordinate system, the x-axis represents the time sequence of sampling points, and the y-axis represents the amplitude of the sampled signal. The encoding process is as follows: Firstly, the time-domain signal is normalized between -1 and 1 in the Cartesian coordinate system.Secondly, encode the value $\tilde{x}_i$ in $\tilde{x}_i$ as the angular cosine ϕ, and encode the timestamp as the radius γ, achieving the conversion from the Cartesian coordinate system to the polar coordinate system. The calculation process is shown in formula (1).

$$\begin{cases} \phi = \arccos \tilde{x}_i; -1 \le \tilde{x}_i \le 1, \tilde{x}_i \in \tilde{X} \\ r = \dfrac{t_i}{N}; t_i \in N \end{cases} \tag{1}$$

Among them, $\tilde{x}_i$ represents the normalized time-series, t_i is the timestamp, and N is a constant factor used to standardize the scale span of the polar coordinate system.

Finally, by comparing the cosine values of the differences in polar angles between each time-series point, the temporal correlations within different time intervals are identified[14]. The calculation process is shown in formula (2)

$$GADF = \begin{bmatrix} \cos(\phi_1 - \phi_1) & \cdots & \cos(\phi_1 - \phi_n) \\ \cos(\phi_2 - \phi_1) & \cdots & \cos(\phi_2 - \phi_n) \\ \vdots & \cos(\phi_i - \phi_i) & \vdots \\ \cos(\phi_n - \phi_1) & \cdots & \cos(\phi_n - \phi_n) \end{bmatrix} \tag{2}$$

After one-dimensional signals are encoded by GADF, a time-series with a length of n is converted into an $n \times n$ matrix. The value of each element represents the grayscale or color intensity of each pixel in the image, reflecting the dynamic patterns and trends of the time-series. The process of converting

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

one-dimensional signals into GADF image data is shown in Figure 2.

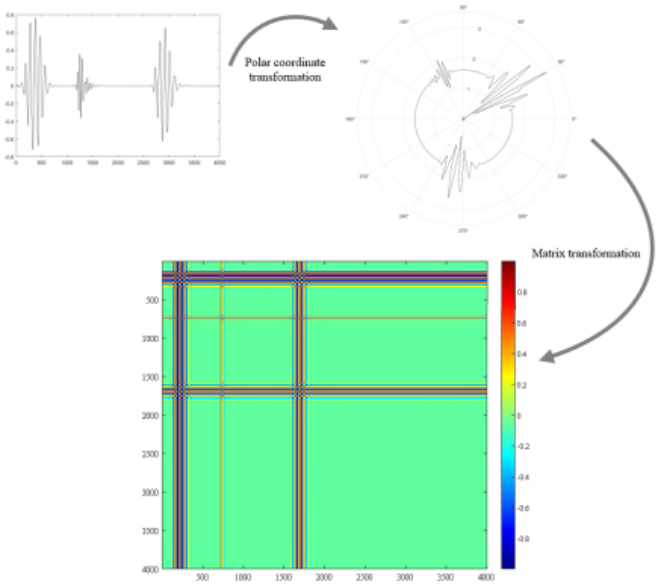


Figure 2. GADF encoding process.

By comparing the cosine values of the polar angle differences between each time series point, the temporal correlations within different time intervals are identified [15]. Each element value represents the grayscale or color intensity of each pixel point in the image, reflecting the dynamic patterns and trends of the time series. The brighter the color (red, orange), it indicates that the relationship between these points is more complex and the variation range is large; the darker the color (blue), it indicates that the relationship between these data points is weaker and the variation is more stable. Based on retaining the temporal characteristics of the signal, the GADF image improves the ability to recognize the nonlinear changes of the heart sound signal by extracting the interdependencies and relationships between the data, which cannot be achieved by traditional spectrograms.

2.2 Vision Transformer (ViT)

In recent years, researchers have introduced the Transformer from the field of natural language processing into computer vision, achieving performance superior to that of the CNN architecture. They named it Vision Transformer (ViT) [16], as shown in Figure 3.

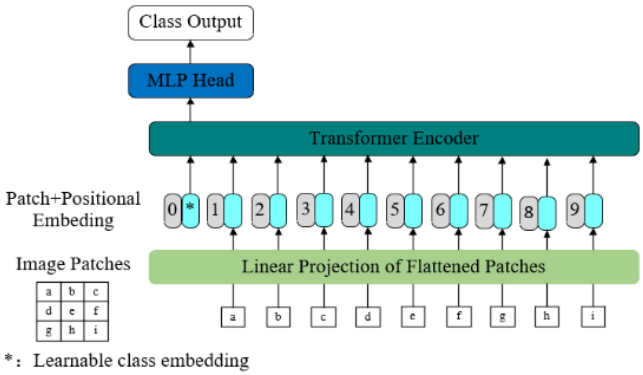


Figure 3. The architecture of ViT

The Transformer Encoder mainly includes a multi-head self-attention layer (MHA) and a multi-layer perceptron block (MLP). Moreover, layer normalization (LN) is performed before both the multi-head self-attention layer and the multi-layer perceptron block. The MLP consists of a fully connected layer, a GELU activation function, and a Dropout layer, which helps the model to capture the non-linear relationships in the data and prevents the model from overfitting due to excessive complexity.

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

2.3 SimAM

SimAM is a lightweight and parameter-free attention mechanism for convolutional neural networks [17]. Without the need to learn additional parameters, it assigns 3D weights for various visual tasks, enhancing the ability of the detection model to extract the features of foreign objects. Its principle is shown in Figure 4.

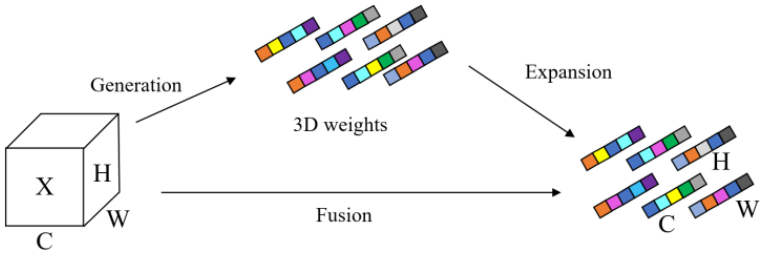


Figure 4. Schematic diagram of the attention mechanism

Among them, C, H and W represent the number of channels, height, and width of the feature map, respectively.

2.4 LeakyReLU

In the recognition of heart sound signals, even a tiny negative fluctuation may contain important physiological information. Therefore, in order to solve the above problems, this paper introduces the LeakyReLU activation function. LeakyReLU (Leaky Rectified Linear Unit) is an improved version of the ReLU (Rectified Linear Unit) activation function [18]. It addresses the problem of "dead neurons" of ReLU in the negative input region by introducing a small slope, thereby enhancing the nonlinear learning ability of the model. The mathematical expression of LeakyReLU is:

$$\text{LeakyReLU}(x) = \max(kx, x) = \begin{cases} x, & \text{if } x > 0 \\ kx, & \text{if } x \le 0 \end{cases}, k \in [0,1] \quad (3)$$

3. Multi-feature Fusion Network Architecture

In this paper, a convolutional neural network (CNN) and a Vision Transformer (ViT) are integrated to construct a multi-feature fusion recognition network for heart sound signals. The network architecture is shown in Figure 5.

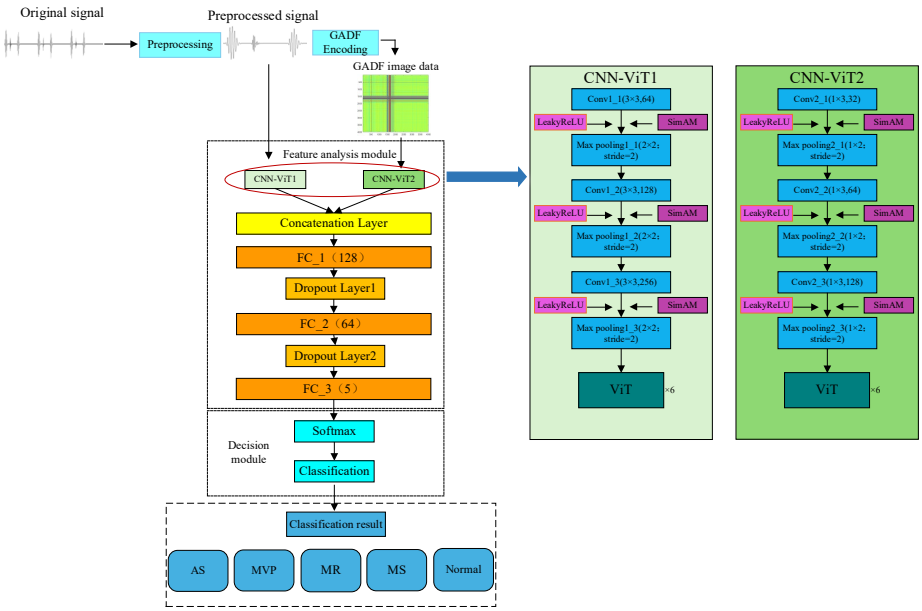


Figure 5. Multi-feature fusion network architecture

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

Firstly, the original heart sound signal is preprocessed and then converted into a GADF image. Subsequently, the one-dimensional signal data is input into CNN-ViT1, and the GADF image data is input into CNN-ViT2. Both types of data undergo local feature extraction through convolutional layers and max pooling layers. On this basis, by adding the SimAM and LeakyReLU activation functions, it helps the model avoid the dead zone problem and capture more important features, especially those small fluctuations that may carry important information. Then the features are input into the ViT module to further mine higher-level semantic information. Through the design of this hybrid model, the advantages of the two different architectures are combined, helping the model better understand the local details in the time-series data as well as the dependencies in the long time series. Finally, the features extracted from the two channels are concatenated according to the channel dimension and input into the fully connected layer. By interspersing Dropout layers, randomly discarding a part of the neurons and their connections, it reduces the model's dependence on specific neurons, ensuring that the model can not only efficiently learn complex features but also maintain good generalization ability and avoid overfitting the training data. The specific parameter configurations of CNN-ViT1 and CNN-ViT2 are shown in Table 1 and Table 2.

Table 1 Parameter of CNN-ViT1

Layers	Parameters
Conv1_1	kernel:3×3, 64; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling1_1	kernel:2×2; stride:2
Conv1_2	kernel:3×3, 128; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling1_2	kernel:2×2; stride:2
Conv1_3	kernel:3×3, 256; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling1_3	kernel:2×2; stride:2

Table 2 Parameter of CNN-ViT2

Structure	Parameters
Conv2_1	kernel:1×3, 32; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling2_1	kernel:1×3; stride:2
Conv2_2	kernel:1×3, 64; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling2_2	kernel:1×2; stride:2
Conv2_3	kernel:1×3, 128; stride:2; padding=1
	SimAM
	LeakyReLU
MaxPooling2_3	kernel:1×2; stride:2

4. Experimental results and analysis

4.1 Data presentation and preprocessing

The heart sound dataset used in this paper includes one normal class (N) and four abnormal classes: aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS), and mitral valve prolapse (MVP). There are 200 signals in each class, making a total of 1000 signals.

According to the heart sound cycle and human physiological characteristics [19], in this paper, the data is segmented with a minimum data length of 2 seconds. Discrete wavelet transform [20-22] is adopted for denoising in this paper. The db8 wavelet basis function is selected, the number of decomposition layers is 10, and the threshold method is the Bayesian threshold method. Finally, since the heart sound frequency of a normal person is approximately 20-600 Hz, according to the Nyquist sampling theorem, all samples are uniformly downsampled to 2 kHz in this paper. The preprocessed signal data contains sufficient information, and there will be no problems such as the data file being too large to process. Each type of signal is shown in Figure 6.

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

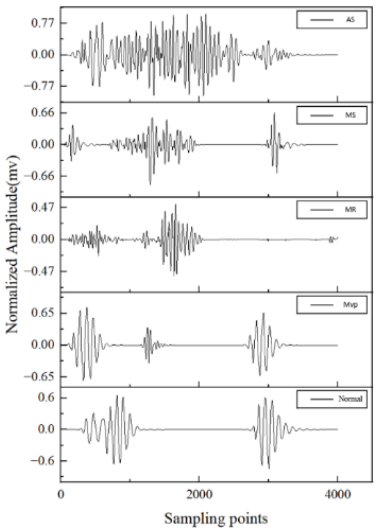


Figure 6. The result of signal preprocessing

It can be clearly seen from the figure that the preprocessing operation reduces the amount of heart sound signal data, and the denoising effect is obvious. It also reduces the computational complexity of subsequent processing, and has no significant impact on the signals within the effective frequency band. Subsequently, digital labels are assigned to each type of heart sound signal, and the details are shown in Table 3.

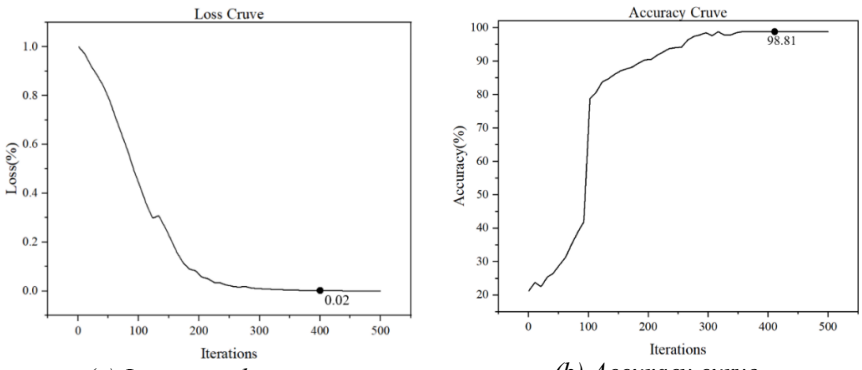
Table 3. DataSet

Species	Sampling rate(kHz)	Sampling points (PCS)	Quantity (PCS)	Label
Normal	2	4000	200	0
AS	2	4000	200	1
MR	2	4000	200	2
MS	2	4000	200	3
MVP	2	4000	200	4

Finally, in this paper, the dataset is divided into a training set, a validation set, and a test set in the ratio of 8:1:1 for the training of the network and the verification of its performance.

4.2 Multi-feature fusion network performance verification

The training set and the validation set are input into the multi-feature fusion network for training. The learning rate is set to 0.001 and is reduced to one-tenth of the original value every 2 epochs. The batch size is 8, the number of epochs is 5, and the optimizer is Adam. The loss function is the standard cross-entropy function, and the total number of iterations is 500. Moreover, the experimental results are evaluated by using accuracy (Acc), sensitivity (Sen), precision (Pre), specificity (Spe), and the comprehensive evaluation index, the F1 score, to assess the classification and recognition performance of the network [23-24]. The curves of the loss rate and the accuracy of the validation set in this paper are recorded, as shown in Figure 7(a) and Figure 7(b).



(a) Loss rate change curve *(b) Accuracy curve*

Figure 7.(a) Loss rate change curve;(b) Accuracy curve

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

In Figure 7(a), when the number of iterations of the validation set data reaches 400, the loss rate is approximately 0.02 and remains in a stable state. At this time, the accuracy in Figure (b) is approximately 98.81%, and the curve also tends to be stable. A confusion matrix is introduced to visualize the prediction results, as shown in Figure 8. The horizontal axis represents the predicted labels, and the vertical axis represents the true labels.

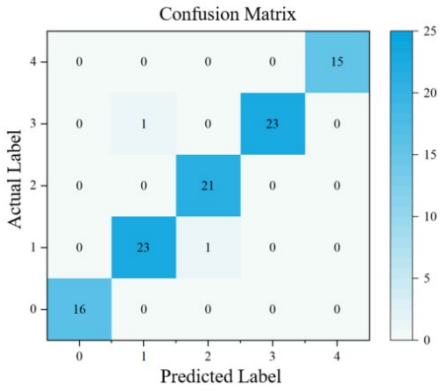


Figure 8.Confusion matrix graph

From the results, it can be seen that the model generally demonstrates excellent classification accuracy and class discrimination ability. The model in this paper can not only effectively distinguish obviously different signal types, but also has good discrimination ability for relatively similar signals.

Finally, in order to further verify the superiority of the method proposed in this paper, the test set is input into another four methods, and a comparison of evaluation indicators is carried out. The results are shown in Table 4.

Table 4 Classification results of different recognition methods for heart sound signals

Method	Acc(%)	Sen(%)	Spe(%)	Pre(%)	F1(%)
SPC-3D[25]	97.49	94.09	98.34	95.90	94.74
PANet[26]	97.89	96.34	98.85	96.96	96.70
Improved CNN-MFCCs[27]	96.59	95.65	97.53	97.47	96.56
TWSVM[28]	94.42	93.65	94.61	95.80	94.78
This paper	98.81	97.41	98.86	98.44	97.98

As can be seen from the table 4, the recognition accuracy of the method in this paper is 98.81%, the sensitivity is 98.41%, the specificity is 97.86%, the precision is 98.84%, and the F1 score is 97.92%. It performs the best on the dataset.

4.3 Comparison experiment between the number of CNN and ViT modules

In this paper, the combination of one convolutional layer and one max-pooling layer is denoted as a convolutional module. In order to determine the specific number of convolutional modules and the number of ViT modules in the network architecture of this paper, this paper conducts 10 experiments on different combinations respectively using the validation dataset, and calculates and records their average accuracy. The results are shown in Figure 9.

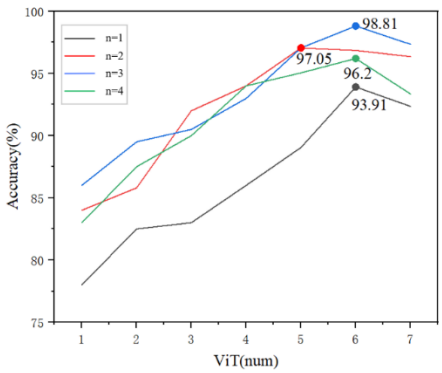


Fig.9 Average accuracy curve

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

Among them, n represents the number of convolutional modules, the horizontal axis represents the number of ViT modules, and the vertical axis represents the recognition accuracy. The four combinations marked in the figure are sequentially recorded as Configuration 1, 2, 3, and 4 according to the accuracy from high to low. It can be seen from the figure that when the number of ViT modules is 6 and the number of convolutional modules is 3, that is, in Configuration 1, the accuracy reaches the highest.

In order to further confirm the selection of the number of modules, this paper records the average time spent in 5 experiments for these 4 configurations, and the results are shown in Table 5.

Table5.Average accuracy and time

Configuration	Average accuracy (%)	time(s)
1	98.81	149.76
2	97.05	141.80
3	96.20	167.41
4	93.91	135.41

Taking both the average accuracy and the time consumption into comprehensive consideration, this paper selects Configuration 1, that is, the combination of 3 convolutional modules and 6 ViT modules, as the construction scheme for the multi-feature fusion network.

4.4 Comparison experiments of different types of graphs

Currently, there are various methods for recognition and classification using graphs or spectra, such as Mel-Frequency Cepstral Coefficients (MFCCs), MTF, and STFT. To verify the superiority of feature fusion between GADF images and one-dimensional signals, this study conducted feature fusion respectively between GADF, MFCCs, MTF, STFT and one-dimensional heart sound signals, and recorded the accuracy curves of the validation set. The results are shown in Figure 10.

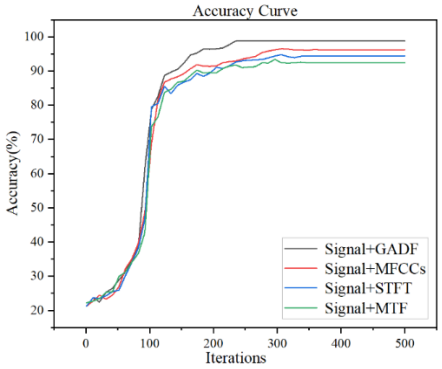


Figure 10. Accuracy curve

As can be seen from the figure 10, when the GADF graph is fused with the one-dimensional signal, the model converges the fastest and achieves the highest final accuracy, converging to 98.81%. Finally, five evaluation metrics, namely Acc, Sen, Spe, Pre, and F1, were used to compare the performance of the four experimental results. The results are shown in Table 6.

Table 6 Recognition and classification results of different types of graphs

Method	Acc(%)	Sen(%)	Spe(%)	Pre(%)	F1(%)
Signal+MTF	92.42	93.09	92.69	92.72	93.06
Signal+STFT	94.41	95.96	96.85	97.82	96.84
Signal+MFCCs	96.21	95.65	97.53	97.50	96.67
Signal+GADF	98.81	97.41	98.86	98.84	98.02

As can be seen from the table 6, GADF performs best when used as the input of the graph, with a final recognition accuracy of 98.81%, a sensitivity of 97.41%, a specificity of 98.86%, a precision of 98.84%, and an F1 score of 98.02%. GADF encodes the temporal sequence and amplitude changes of one-dimensional signals through the method of angular difference. This encoding method can capture the dynamic characteristics of both time and amplitude simultaneously, and more effectively reveal the nonlinear relationships existing within the signals.

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

5. Conclusion

In this paper, GADF images and one-dimensional signals are used as inputs, and a multi-feature fusion network for heart sound signals is constructed by integrating a Convolutional Neural Network (CNN) and a Vision Transformer (ViT). First, the one-dimensional signals and GADF image data are respectively input into the CNN-ViT1 and CNN-ViT2 channels for feature extraction. By combining the SimAM module and the LeakyReLU activation function, the network avoids the problem of "dead neurons" and, at the same time, enhances the model's non-linear learning ability and improves the network's recognition ability for key signal features. Then, the features extracted from the two channels are concatenated and fused in the channel dimension, enabling the model to obtain more comprehensive signal features and improving the recognition accuracy of heart sound signals. To verify the performance of the model, a 5-classification dataset is used in this paper. Through multiple comparative experiments, it is verified that the method proposed in this paper has the best overall performance. The experimental results show that the recognition accuracy of the method proposed in this paper is 98.81%, the sensitivity is 98.41%, the specificity is 97.86%, the precision is 98.84%, and the F1-score is 97.92%, demonstrating the superiority of the method.

Acknowledgements

The author would like to express gratitude to all the authors cited in this paper and the financial support from the Free Exploration Project of the Natural Science Foundation of Jilin Province.

References

[1] Griffel B, Zia M K, Fridman V, et al. Path length entropy analysis of diastolic heart sounds[J]. Computers in biology and medicine, 2013, 43(9): 1154-1166.
[2] Wang Y, Li W, Zhou J, et al. Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on OMS-WPD[J]. Future Generation Computer Systems, 2014, 37: 488-495.
[3] Zheng Y, Guo X, Ding X. A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification[J]. Expert Systems with Applications, 2015, 42(5): 2710-2721.
[4] Deng S W, Han J Q. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps[J]. Future Generation Computer Systems, 2016, 60: 13-21.
[5] Li P, Ge J, Li H. Lysine acetyltransferases and lysine deacetylases as targets for cardiovascular disease[J]. Nature Reviews Cardiology, 2020, 17(2): 96-115.
[6] Roth G A, Mensah G A, Johnson C O, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study[J]. Journal of the American college of cardiology, 2020, 76(25): 2982-3021.
[7] Tsao C W, Aday A W, Almarzooq Z I, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association[J]. Circulation, 2022, 145(8): e153-e639.
[8] Milani M G M, Abas P E, De Silva L C, et al. Abnormal heart sound classification using phonocardiography signals[J]. Smart Health, 2021, 21: 100194.
[9] Mei N, Wang H, Zhang Y, et al. Classification of heart sounds based on quality assessment and wavelet scattering transform[J]. Computers in biology and medicine, 2021, 137: 104814.
[10] Chen J, Guo Z, Xu X, et al. A robust deep learning framework based on spectrograms for heart sound classification[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023.
[11] Wang Y, Yang X, Qian X, et al. Assistive diagnostic technology for congenital heart disease based on fusion features and deep learning[J]. Frontiers in Physiology, 2023, 14: 1310434.
[12] Zhang H, Zhang P, Wang Z, et al. Multi-feature decision fusion network for heart sound abnormality detection and classification[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 28(3): 1386-1397.
[13] Zhang X, Liu X, Liu G. A Heart Sound Signal Classification Method Based on the Mixed Characteristics of Mel Cepstrum Coefficient and Second-Order Spectrum[J]. Circuits, Systems, and Signal Processing, 2024, 43(6): 3533-3552.
[14] Wang Z , Oates T .Imaging Time-Series to Improve Classification and Imputation //International Conference on Artificial Intelligence. AAAI Press, 2015: 3308-4132.
[15] Wang Z, Oates T. Imaging time-series to improve classification and imputation[J]. arXiv preprint arXiv:1506.00327, 2015.
[16] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for

ISSN 2616-5775 Vol. 8, Issue 4: 1-10, DOI: 10.25236/AJCIS.2025.080401

image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
[17] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR, 2021: 11863-11874.
[18] Maniatopoulos A, Mitianoudis N. Learnable leaky relu (LeLeLU): An alternative accuracy-optimized activation function[J]. Information, 2021, 12(12): 513.
[19] Li S, Li F, Tang S, et al. A review of computer-aided heart sound detection techniques[J]. BioMed research international, 2020, 2020(1): 5846191.
[20] Chen Y, Wei S, Zhang Y. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network[J]. Medical & Biological Engineering & Computing, 2020, 58: 2039-2047.
[21] Chen P, Zhang Q. Classification of heart sounds using discrete time-frequency energy feature based on S transform and the wavelet threshold denoising[J]. Biomedical Signal Processing and Control, 2020, 57: 101684.
[22] Varghees V N, Ramachandran K I. Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope[J]. IEEE Sensors Journal, 2017, 17(12): 3861-3872.
[23] Yingwei L, Sundararajan N, Saratchandran P. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm[J]. IEEE Transactions on neural networks, 1998, 9(2): 308-318.
[24] Augusteijn M F, Clemens L E, Shaw K A. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier[J]. IEEE Transactions on Geoscience and Remote Sensing, 1995, 33(3): 616-626.
[25] Zhou X, Guo X, Zheng Y, et al. Detection of coronary heart disease based on MFCC characteristics of heart sound[J]. Applied Acoustics, 2023, 212: 109583.
[26] Deng E, Jia Y, Zhu G, et al. Heart sound signals classification with image conversion employed[J]. Electronics, 2024, 13(7): 1179.
[27] Wang J J, Xiong F L. Research on the recognition method of heart sound signal based on improved MFCC and CNN. Computer Measurement and Control, 2024, 32(12): 201-207+215.
[28] Ismail Fawaz H, Lucas B, Forestier G, et al. Inceptiontime: Finding alexnet for time series classification[J]. Data Mining and Knowledge Discovery, 2020, 34(6): 1936-1962.