



OPEN ACCESS

EDITED BY

Pritpal Singh,
Central University of Rajasthan, India

REVIEWED BY

Anoop Tiwari,
Central University of Haryana, India
Subodh Kumar,
Central University of Rajasthan, India

*CORRESPONDENCE

Fan Fan
✉ fanfan_bch@163.com

RECEIVED 18 November 2024

ACCEPTED 15 January 2025

PUBLISHED 29 January 2025

CITATION

Zeng P, Kang S, Fan F and Liu J (2025)
Enhanced heart sound anomaly detection via
WCOS: a semi-supervised framework
integrating wavelet, autoencoder and SVM.
Front. Neuroinform. 19:1530047.
doi: 10.3389/fninf.2025.1530047

COPYRIGHT

© 2025 Zeng, Kang, Fan and Liu. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Enhanced heart sound anomaly detection via WCOS: a semi-supervised framework integrating wavelet, autoencoder and SVM

Peipei Zeng¹, Shuimiao Kang^{2,3}, Fan Fan^{4*} and Jiyuan Liu^{2,5}

¹Civil Aviation University of China Engineering Technology Training Center, Civil Aviation University of China, Tianjin, China, ²Aviation Ground Special Equipment Research Base of Civil Aviation Administration of China, Civil Aviation University of China, Tianjin, China, ³College of Safety Science and Engineering, Civil Aviation University of China, Tianjin, China, ⁴Pediatric Cardiac Center, Beijing Children's Hospital, Capital Medical University, Beijing, China, ⁵College of Aeronautical Engineering, Civil Aviation University of China, Tianjin, China

Anomaly detection is a typical binary classification problem under the condition of unbalanced samples, which has been widely used in various fields of data mining. For example, it can help detect heart murmurs when the heart is structurally abnormal, to tell if a newborn has congenital heart disease. Due to the low time and high efficiency, most work focuses on the semi-supervised anomaly detection method. However, the anomaly detection effect of this method is not high because of massive data with uneven samples and different noise. To improve the accuracy of anomaly detection under unbalanced sample conditions, we propose a new semi-supervised anomaly detection method (WCOS) based on semi-supervised clustering, which combines wavelet reconstruction, convolutional autoencoder, and one classification support vector machine. In this way, we can not only distinguish a small proportion of abnormal heart sounds in the huge data scale but also filter the noise through the noise reduction network, thus significantly improving the detection accuracy. In addition, we evaluated our method using real datasets. When the noise of $\sigma = 0.5$, the AUC standard deviation of the WR-CAE-OCSVM is 19.2, 54.1, and 29.8% lower than that of WR-OCSVM, CAE-OCSVM and OCSVM, respectively. The results confirmed the higher accuracy of anomaly detection in WCOS compared to other state-of-the-art methods.

KEYWORDS

heart sound detection, semi-supervised anomaly detection, sample imbalance, convolutional autoencoder, one classification support vector machine

1 Introduction

With the change in the modern medical model, the spectrum of human diseases and death has undergone great changes, and birth defects have gradually become the main cause threatening children's health, congenital heart disease is the most common type of birth defect disease, accounting for about 28% of all congenital malformations. Among the fatal defects in children under 5 years old, congenital heart disease is the first (Xue et al., 2022). Therefore, early and accurate diagnosis of congenital heart disease will make children get timely diagnosis and treatment, and significantly improve the prognosis of children. Cardiac auscultation, as a convenient and non-invasive examination method, is the most important means of early screening for congenital heart disease, and the accuracy of the results of auscultation has become an important factor

affecting the screening effect of congenital heart disease. The traditional auscultation method requires the use of a general stethoscope by an audiologist with certain audiological skills and experience, but because of the high degree of subjectivity, the lack of audiological skills and experience of primary care doctors is often a bottleneck limiting the effectiveness of congenital heart disease screening. With the development of artificial intelligence technology, the collection of heart sounds by electronic stethoscope and the recognition of digital auscultation data by artificial intelligence algorithm make the artificial intelligence of cardiac auscultation possible (Wang et al., 2024).

Artificial intelligence auscultation will effectively assist doctors in judging the condition, greatly reduce the work intensity of doctors, and improve the accuracy of auscultation. Heart sound is a complex sound produced by the switching of heart valves, the relaxation and contraction of tendons and muscles, the impact of blood flow, and the vibration of the cardiovascular wall. Abnormal blood shunt will occur when the heart structure is abnormal, and then the heart murmur will be generated. Therefore, the detection of abnormal heart sounds is becoming an increasingly important research field (Jyothi and Pradeepini, 2021).

Abnormal detection will have different degrees of detection difficulties according to different samples and different processes of signal acquisition. First, for heart sound samples, sample imbalance will cause the model to be biased toward most normal samples, thus placing a higher emphasis on most normal samples in prediction, which may lead to the improvement of the accuracy of most normal samples and the decrease of the accuracy of a few abnormal samples. This bias makes it impossible for the model to accurately distinguish abnormal heart sound samples because abnormal samples usually belong to a small number of abnormal samples (Liang et al., 2021). At the same time, the heart sound signal is easily interfered with by various noises in the acquisition process, such as power line frequency noise, baseline drift, myoelectric interference (Xiaodong et al., 2020; Shen et al., 2021), etc. These noises will reduce the quality of the heart sound signal and lead to the loss of information, which will affect the accuracy of the subsequent heart sound signal analysis and processing. Consequently, sample imbalance will cause the model to bias most normal classes and reduce the prediction accuracy of a few abnormal classes, while noise interference will reduce the robustness and feature learning ability of the model.

Effective identification of normal and abnormal heart sounds is a difficulty in artificial intelligence auscultation research (Ling et al., 2003), and it is urgent to establish an anomaly detection technology with a better detection effect to reduce the influence of sample heterogeneity and noise on detecting abnormal heart sounds. Deep learning technology has made a lot of progress in the field of anomaly detection, which involves a variety of deep neural network structures, loss functions, and optimization algorithms (Chong et al., 2021), including anomaly detection based on autoencoders, anomaly detection based on graph neural networks and anomaly detection based on deep generation models. Below is a detailed summary of the methods for detecting heart sound anomalies based on deep learning techniques.

2 Literature review of anomaly detection in heart sounds

The autoencoder is an unsupervised deep neural network that learns a compressed representation from the input data and reconstructs an output that is as similar as possible to the original data. Abnormal

data usually cannot be reconstructed well, so the autoencoder can be used for anomaly detection. Recent studies have also combined other deep-learning techniques with autoencoders, such as autoencoders (Chen et al., 2024), variational autoencoders (Gangloff et al., 2024), and generative adversarial networks (Ruff et al., 2021; Li, 2024). The disadvantage is that the autoencoder may learn abnormal data features and thus reduce the detection accuracy, and the reconstruction error cannot accurately reflect the abnormal degree of the data.

Graph neural network is a kind of deep learning model specially used to process graph-structured data (Mir et al., 2023). Abnormal data is usually some unusual node or edge in the data, so you can translate the anomaly detection problem into detecting abnormal nodes or edges in the graph structure. The latest research shows that the anomaly detection method based on graph neural networks has achieved very good results in the heart sound signal recognition scene (Rezaee et al., 2022) and a variety of other scenes (Niu et al., 2021). To alleviate the nature and scalability of the scene, combined with other methods, variational graph convolutional networks (Mir et al., 2024), two-domain graph convolutional networks (Li et al., 2023), space-time graph networks (Yuan et al., 2024), and so on are proposed. The disadvantage is that the computational complexity of the graph neural network is high, which requires more computational resources, and the model is highly dependent on the accuracy and integrity of the graph structure, which affects the anomaly detection effect.

Deep generation models are a class of deep learning models that can learn a probability distribution from data and generate new data similar to the original data. The latest research shows that the anomaly detection method based on the deep generation model has high flexibility and robustness, and can be applied to various types of data, such as text data (Fan et al., 2021), image data (Sanders et al., 2020), and time series data (Gao et al., 2021).

There are also other methods based on mathematical models, such as establishing an indiscernibility-assisted intuitionistic fuzzy-rough set model based on fuzzy and rough set theories to reduce the noise (Shreevastava et al., 2023) establishing a missing value estimation and feature selection method to reduce the dimensionality while maintaining the performance (Jain et al., 2023), integrating the fuzzification module and the RBFNN, designing an adaptive control scheme to effectively reduce the model's detection uncertainty (Kumar et al., 2024), and designing an adaptive control scheme based on intuitionistic fuzzy interference to resist noise and better handle uncertainty in judgment and recognition (Jain et al., 2022). The shortcomings of the mathematical model-based approach are that the model efficacy depends on the data quality and distribution, the parameter tuning is challenging, and the generalization to different datasets needs to be further investigated.

To sum up, anomaly detection based on autoencoders performs well in terms of simplicity and generalization ability but lacks in terms of sensitivity to abnormal data and reconstruction error for heart sound data (Cloudera and Nisha, 2020). Anomaly detection based on graph neural networks performs well in handling complex relationships and unifying frameworks, but it is highly computation-based and depends on graph structure (Jia-Yan et al., 2020). However, anomaly detection based on a depth generation model has significant advantages in terms of representation ability, automatic feature extraction, and wide applicability of cardiac sound anomaly data (Xing et al., 2020). Therefore, based on the deep generation model and inspired by the DCGAN network architecture model (Liu et al., 2019) and the semi-supervised value detection framework (Shi et al., 2023),

this study intends to collect the heart sound data of children in our center. Combined with the semi-supervised anomaly detection framework of wavelet reconstruction (Aziz et al., 2024), convolutional autoencoder (Li et al., 2024), and one classification support vector machine (Zhang et al., 2025), a method for differential analysis of normal heart sounds and abnormal heart sounds was constructed and verified, the name of the framework is WCOS, which is short for WR-CAE-OCSVM. The research results will provide a reference for further research on the classification of heart noises.

The subsequent chapters of this paper are arranged as follows. Firstly, Chapter 2 describes the framework construction process, focusing on wavelet reconstruction and convolutional autoencoder. Secondly, as an application verification, Chapter 3 uses real data to verify the validity of the proposed model. Finally, in Chapter 4, we give the main conclusions of the paper and future research work.

The subsequent chapters of this paper are organized as follows. Firstly, the opening of Chapter 2 describes the key significance of heart sound abnormality detection in the diagnosis of congenital heart disease and puts forward the difficulties of sample imbalance and noise interference, followed by Chapter 3, which comprehensively analyzes the limitations of traditional means and the advantages and disadvantages of each method of deep learning, and analyzes the performances of a variety of models, and Chapter 4, which describes the process of constructing the framework of the WCOS model and analyzes the principles of noise reduction by wavelet, convolutional self-encoder feature mining, and classification by support vector machine. In Chapter 5, the validity of the proposed model is verified based on heart sound signals collected by an electronic stethoscope. Finally, in Chapter 6, we give the main conclusions of the article as well as an analysis of the model's strengths and weaknesses.

3 Theory of the developed method

In this paper, we organically combine the wavelet reconstruction (WR), the convolutional auto-encoder (CAE), and one classification

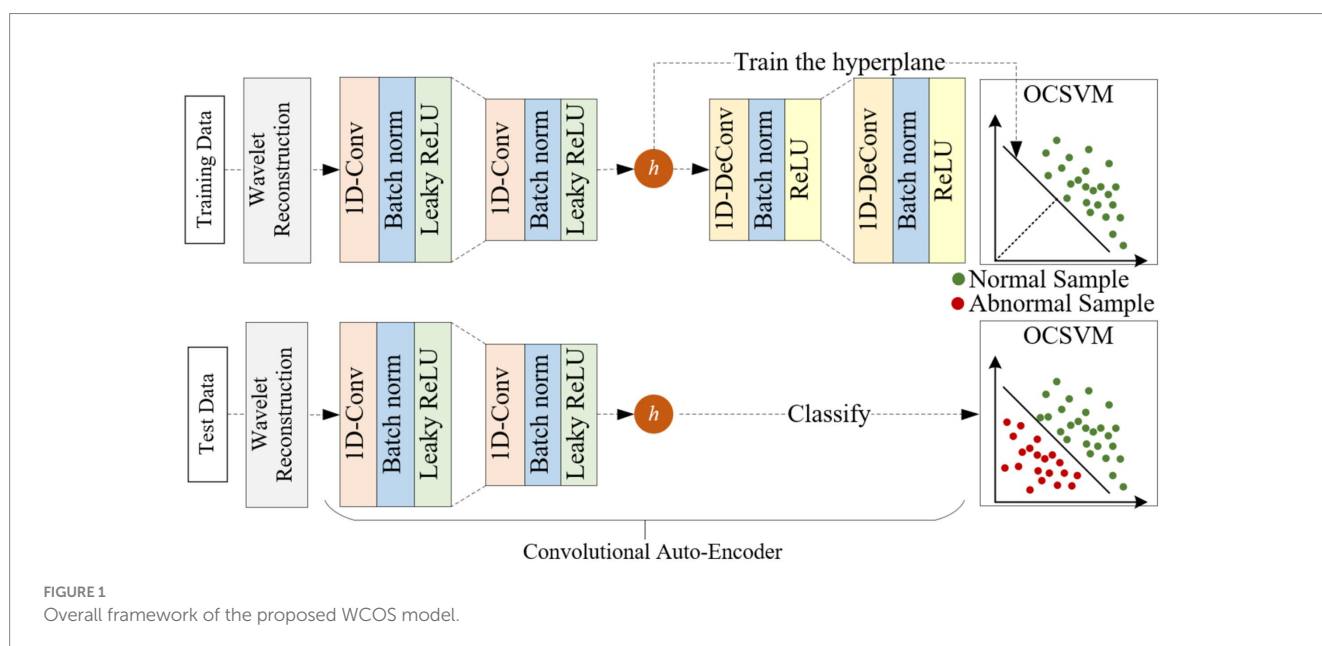
support vector machine (OCSVM) to construct a new semi-supervised exception detection framework. This section first introduces the overall process of the proposed framework and briefly introduces the training and testing process of the framework. Then, we introduce the important components of the framework: wavelet reconstruction, convolutional autoencoder, and one classification support vector machine.

3.1 Fundamental architectures of the WCOS model

The overall framework of WCOS proposed in this paper mainly includes three parts: wavelet reconstruction, convolutional autoencoder, and one classification support vector machine, as shown in Figure 1.

The upper part of Figure 1 shows the training process of the WCOS model. Firstly, the original signal is decomposed by a multilayer wavelet, the high-frequency information containing noise is zeroed, and then the low-frequency information is reconstructed to obtain the reconstructed signal. Then, the reconstructed signals are divided by the sliding window method to obtain one sample after another. Finally, the training set and the test set are divided by the method of five-fold crossover. After partitioning the data set, the WCOS model is trained. The training steps of the model are summarized as follows: (A) Train the convolutional autoencoder using normal samples, the loss function – reconstruction error is minimized by backpropagation algorithm; (B) Train the one classification support vector machine using the potential representation of normal samples, minimize the loss function.

The bottom section of Figure 1 shows the testing process for the WCOS model. The testing steps of the model are summarized as follows: (A) Input the test sample into the convolutional autoencoder and obtain the corresponding latent representation; (B) Input the potential representation of the sample into the one classification support vector machine and obtain the corresponding anomaly score.



Finally, the abnormal score was used to diagnose patients with cardiac abnormalities.

Next, the important components of the model are described in detail: wavelet reconstruction, convolutional self-encoder, and one-classification support vector machine.

3.2 Wavelet reconstruction

Wavelet reconstruction is used to denoise heart sounds. As a classical time-scale analysis algorithm, wavelet transform can decompose the signal into two sets of wavelet coefficients: approximation coefficient and detail coefficient. As shown in Figure 2, the approximate coefficients can be decomposed recursively, allowing for a more “detailed” examination of the original signal. In general, the approximate coefficient represents the information of the low-frequency part of the signal and is considered as the trend term of the signal. The detail coefficient characterizes the information of the high-frequency part of the signal, which is considered as the noise term of the signal. Therefore, to avoid the influence of noise, it is necessary to retain the approximate coefficient and discard the detail coefficient in the process of reconstruction.

In practical engineering, a series of low-pass filters and high-pass filters are usually used for discrete wavelet transform. Specifically, the wavelet decomposition formula for the original signal x is shown as follows function of Equation 1:

$$\begin{cases} cA(k) = \sum x(n)h(2k - n) \\ cD(k) = \sum x(n)g(2k - n) \end{cases} \quad (1)$$

where $x(n)$ denotes the sampled value of the original signal at discrete moments, $cA(k)$ denotes the approximation coefficient of the signal, $cD(k)$ denotes the detail coefficient, $h(\cdot)$ denotes the low-pass filter, $g(\cdot)$ denotes the high-pass filter, k is the index of the coefficients after the discrete wavelet transform, n is the sampling index of the original signal.

Wavelet reconstruction is the inverse operation of wavelet decomposition, and the formula of wavelet reconstruction is as follows function of Equation 2:

$$x(k) = \sum_n cA(n)h(k - 2n) + \sum_n cD(n)g(k - 2n) \quad (2)$$

Where $x(k)$ denotes the sampled value of the reconstructed signal at discrete moments, $cA(n)$, $cD(n)$. Same meaning as in wavelet decomposition, approximation coefficients and detail coefficients, respectively, but here as inputs to the reconstruction, $(k - 2n)$, $g(k - 2n)$ still represent low-pass and high-pass filters, respectively, but the indexing is different here and is the operation used to reconstruct the original signal from the coefficients.

To avoid noise, set the detail parameters $cD(n)$ of the above equation to zero, then the above formula becomes the function of Equation 3 below:

$$x(k) = \sum_n cA(n)h(k - 2n) \quad (3)$$

Where, $cA(n)$ denotes the approximation coefficient of the signal. This removes the high-frequency noise component of the signal, which tends to contain more noise, while the low-frequency component retains the main information of the signal. At this time, the reconstructed signal not only retains the main information of the original signal but also avoids the influence of noise to a certain extent, so the reconstructed signal is more suitable for the detection of cardiac abnormalities.

The Symlet 4 (sym4) wavelet was used in this study for wavelet decomposition. The sym4 wavelet was chosen because it provides a favorable trade-off between time and frequency localization of heart sound signals. The complexity of heart sound waveforms, which are often disturbed by noise, calls for a wavelet that can accurately decompose both low-frequency fundamental rhythms as well as high-frequency transient events such as murmurs. sym4 wavelet’s design features make it capable of such tasks. Its symmetrical nature reduces phase distortion during decomposition and reconstruction, which is essential for maintaining the integrity of the signal’s phase information, which is critical for recognizing subtle differences in heartbeat patterns.

Regarding the level of wavelet decomposition, a 5-level decomposition was used. The five-level decomposition provides a fine-grained exploration of the signal spectrum. The initial levels capture the broad low-frequency trends that underpin the normal cardiac cycle. As the level of decomposition increases, details and potential anomalies in the high-frequency range gradually become apparent. For example, subtle changes in murmur intensity and frequency can be better separated and characterized at these higher levels. Thus, this five-level approach allows the extraction of discriminative features from different frequency layers and improves the efficiency of subsequent anomaly detection procedures by providing a comprehensive spectral depiction of the heart sound signal.

3.3 1-D convolutional auto-encoder

Convolutional autoencoders are used to extract the characteristic information of heart sounds. The convolutional autoencoder consists of an encoder and a decoder. The special feature is that the convolutional autoencoder uses a convolutional layer to replace the fully connected layer in the encoder, and a deconvolution layer to

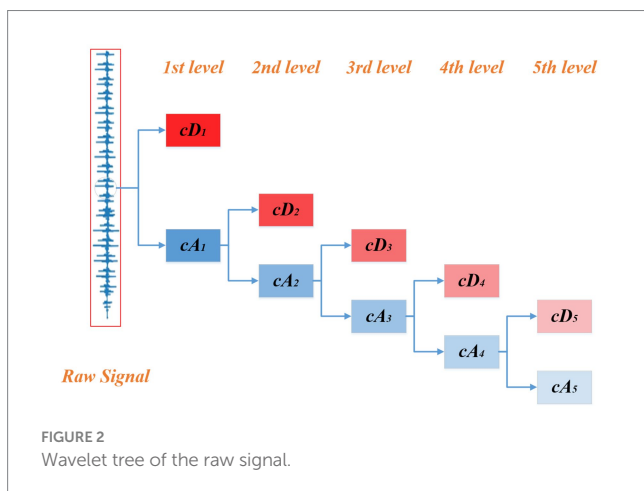


FIGURE 2 Wavelet tree of the raw signal.

replace the fully connected layer in the decoder. The specific structure of the convolutional autoencoder in this paper is shown in Figure 1: the encoder consists of two convolutional layers, a batch layer, and a Leaky ReLU activation function, while the decoder consists of two deconvolution layers, a batch layer, and a ReLU activation function, and finally a Tanh layer.

The convolutional layer is an important component of convolutional autoencoders, which can greatly reduce the number of parameters in the network. On the one hand, it can improve the robustness of the network; On the other hand, it can also reduce the risk of network overfitting. For the single-channel input x , the potential representation of the i feature map is formally defined as follows Equation 4:

$$h^i = f(x \otimes W^i + B^i) \tag{4}$$

Where f represents the activation function in the encoder (we use Leaky ReLU in this article), B^i represents the bias of the i feature map, and the symbol \otimes represents the two-dimensional convolution operation.

The deconvolution layer can be seen as an inverse mapping of the convolution layer, and the reconstructed representation for the input x is defined as follows Equation 5:

$$\hat{x} = g\left(\sum_{i \in H} h^i \otimes \hat{W}^i + c\right) \tag{5}$$

Where g represents the activation function in the decoder (ReLU is used in this article), c represents the bias of each input channel, and H represents a set of potential feature maps.

In addition, the loss function of the convolutional autoencoder is usually the reconstruction error of the sample, which can be expressed as follows Equation 6 below:

$$L_{cae} = \frac{1}{|x|} \sum_{x \in X} \|x - \hat{x}\|_1 \tag{6}$$

Where the $\|\cdot\|_1$ represents the 1-norm. 1-norm can obtain a clearer reconstructed sample than 2-norm, so this paper uses 1-norm to measure the reconstruction error between the input sample and the reconstructed sample.

3.4 One-class support vector machine

OCSVM is used to diagnose heart sounds based on characteristic information. After the potential characteristics of samples are obtained by CAE, the boundary of normal samples needs to be learned to distinguish normal samples from abnormal samples. As a classical semi-supervised learning algorithm, the one classification SVM only uses normal samples for training, so compared with supervised algorithms, it can avoid the training problems caused by the high imbalance between normal samples and abnormal samples. As shown in Figure 3, the basic principle of OCSVM is to map the normal sample into a high-dimensional space through the kernel function,

treat the origin as the only outlier, learn an optimal hyperplane (with maximum spacing between the origin and the normal sample), and can distinguish the origin from the normal sample.

For a given data set $H = \{h_i : i = 1, 2, \dots, N\}$ and the features of the data mapped to high-dimensional space mapping function $\varphi(\cdot)$, where $\varphi(\cdot)$ can be computed by the kernel function of Equation 7:

$$k(h_i, h_j) = \varphi(h_i)^T \varphi(h_j) \tag{7}$$

where h_i and h_j are two data points and $\varphi(\cdot)$ is the feature mapping function. To make data set H away from the origin, equivalent to the following optimization problem Equation 8:

$$\begin{aligned} \min_{w, \xi, v} & \frac{1}{2} \|w\|^2 + \frac{1}{CN} \sum_i \xi_i - v \\ \text{subject to} & \begin{cases} (w \cdot \varphi(h_i)) \geq v - \xi_i \\ \xi_i \geq 0 \end{cases} \end{aligned} \tag{8}$$

where w represents a vector perpendicular to the hyperplane; ξ_i represents the relaxation variable, which can solve the outlier in the training set. v is the distance to the origin; $C \in (0, 1]$ represents the control factor that controls the complexity of the model, N is the total number of samples in the dataset H , subject to denote the constraints, and $\varphi(\cdot)$ is the previous kernel function formula. It is worth noting that in this paper the dataset H consists of potential representations of samples.

After solving the above optimization problem, to obtain the weight of a set of models $\{w_i : i = 1, 2, \dots, N\}$, for any of the test sample h_p , its corresponding decision function is Equation 9 below:

$$f(h_p) = \text{sgn}\left(\sum_i w_i k(h_p, h_i) - v\right) \tag{9}$$

Where $\text{sgn}(\cdot)$ represents a symbolic function. If the decision function $f(h_p) = -1$ of sample h_p is tested, then sample h_p is diagnosed as normal. Conversely, if the decision function of test sample h_p , then test sample h_p is diagnosed as an abnormal sample. Thus anomaly detection under unbalanced conditions can be realized.

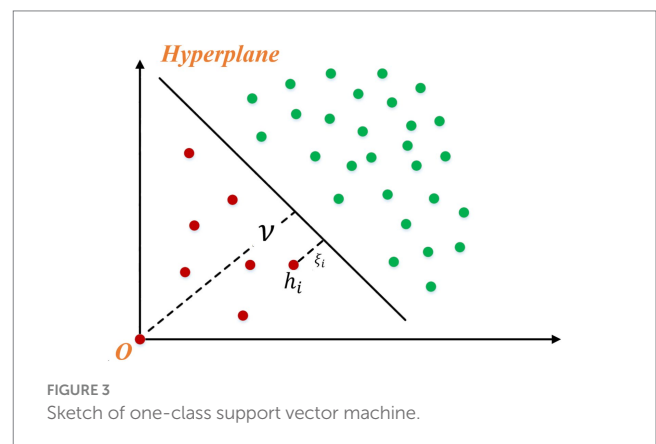


FIGURE 3 Sketch of one-class support vector machine.

3.5 Evaluation metrics of the WCOS model

The Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC) was used as the evaluation index, ROC abscissa is the false positive rate (FPR), the ordinate is the true positive rate (TPR), and the area surrounded by the coordinate axis is defined as AUC. Studies show that AUC can better measure the performance of the classifier than the overall accuracy under the condition of uneven data. AUC is an index to evaluate the quality of the binary classification model, given by Equation 10:

$$AUC = \frac{\sum_{i \in positionClass} rank_i - \frac{M(1+M)}{2}}{M \times N} \quad (10)$$

$$\sum_{i \in positionClass} rank_i$$

where *positionClass* represents the positive set, *rank_i* represents the rank of the *i*th sample in the sample ranking, and the $\sum_{i \in positionClass} rank_i$ term represents the sum of the ranks belonging to the positive samples. M and N represent True Positive Rate (TPR) and False Positive Rate (FPR), respectively.

4 Experimental results

4.1 Data set preparation and preprocessing

To verify the validity of the constructed heart sound abnormality detection model, a Littmann 3200 electronic stethoscope from 3 M Company, USA, was used to collect heart sound signals to construct the dataset. The data were processed using a 5-fold cross-validation method, and the data were derived from actual clinical measurements in multiple medical institutions, not from simulations. In clinical practice, abnormalities in cardiac structure or function are associated with changes in the rhythm, frequency, and intensity of heart sounds. For example, murmurs caused by valvular lesions and weakened or enhanced heart sounds due to myocardial lesions are selected as key indicators for the detection of abnormalities. At the same time, we searched patients' medical records, diagnostic reports, and treatment feedback to accurately determine the status of heart sounds, constructed samples using sliding window technology, and analyzed a large number of cases to clarify that the length of abnormal heart sound sequences was mostly in the range of 6 to 10, and then set the sliding window size to 10.

This paper is experimentally verified on a device with a six-core Intel(R) Core (TM)i7-9750H CPU@2.59GHz processor and 8GB DDR4 memory.

4.2 Experimental setup

The original signal statistics method TF24 calculates the original signal through a series of specific formulas, and its role is to decompose the original sound signal into 24 parameters, which cover a variety of characteristics of the signal in the time domain and frequency domain. The main purpose of TF24 is to extract representative features from the original signal, which can describe

the characteristics of the original signal more comprehensively. The usage of TF24 is to use its 24 extracted parameters as features as the input of subsequent models (such as OCSVM). The time domain parameters are expressed as p1 ~ p11 and given by Equation 11:

$$\left\{ \begin{array}{ll} p_1 = \frac{\sum_{n=1}^N x(n)}{N} & p_2 = \sqrt{\frac{\sum_{n=1}^N (x(n) - p_1)^2}{N-1}} \\ p_3 = \left(\frac{\sum_{n=1}^N \sqrt{|x(n)|}}{N} \right)^2 & p_4 = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}} \\ p_5 = \max |x(n)| & p_6 = \frac{\sum_{n=1}^N (x(n) - p_1)^3}{(N-1)p_2^3} \\ p_7 = \frac{\sum_{n=1}^N (x(n) - p_1)^4}{(N-1)p_2^4} & p_8 = \frac{p_5}{p_4} \\ p_9 = \frac{p_5}{p_3} & p_{10} = \frac{p_4}{\frac{1}{N} \sum_{n=1}^N |x(n)|} \\ p_{11} = \frac{p_5}{\frac{1}{N} \sum_{n=1}^N |x(n)|} & \end{array} \right. \quad (11)$$

In Equation 11, p1 represents the average amplitude of the signal in the time domain, p2 calculates the standard deviation of the signal in the time domain, p3 is related to some weighted summation of the signal amplitude, p4 represents the root mean square value of the signal in the time domain, p5 directly obtains the maximum amplitude of the signal in the time domain, p6–p11 These parameters involve higher order statistics of the signal amplitude relative to the mean.

In addition, the original signal statistical method decomposed the sound signal into 13 parameters in the time domain, expressed as p12 ~ p24, which is given by Equation 12:

$$\left\{ \begin{array}{ll} p_{12} = \frac{\sum_{k=1}^K s(k)}{K} & p_{13} = \frac{\sum_{k=1}^K (s(k) - p_{12})^2}{K} \\ p_{14} = \frac{\sum_{k=1}^K (s(k) - p_{12})^3}{K(\sqrt{p_{13}})^3} & p_{15} = \frac{\sum_{k=1}^K (s(k) - p_{12})^4}{Kp_{13}^2} \\ p_{16} = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)} & p_{17} = \sqrt{\frac{\sum_{k=1}^K (f_k - p_{16})^3 s(k)}{K}} \\ p_{18} = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}} & p_{19} = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}} \\ p_{20} = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K f_k^4 s(k) \sum_{k=1}^K s(k)}} & p_{21} = \frac{p_{17}}{p_{16}} \\ p_{22} = \frac{\sum_{k=1}^K (f_k - p_{16}) s(k)}{Kp_{17}^2} & p_{23} = \frac{p_{16}}{\sum_{k=1}^K \sqrt{f_k - p_{16}} s(k)} \\ p_{24} = \frac{\sum_{k=1}^K (f_k - p_{16})^4 s(k)}{K\sqrt{p_{17}}} & \end{array} \right. \quad (12)$$

In Equation 12, p12 represents the vibration energy in the frequency domain, and p13-p15, p17, and p21-p24 represent the convergence of the power spectrum. p16, p18, p19 and p20 represents the pattern of change of the main frequency.

The experiments involved four modules, namely OCSVM model, raw signal statistics TF24, convolutional autoencoder, and wavelet decomposition. Five models were set up for comparison, the control group is shown in Table 1.

As can be seen in the above table, the first group OCSVM represents a classification support vector machine, the second group TF24-OCSVM represents the combination of raw signal statistical method (TF24) and OCSVM, and the third group CAE-OCSVM represents the combination of convolutional autoencoder (CAE) and OCSVM. The fourth group of WR-OCSVM represents the combination of wavelet reconstruction (WR) and OCSVM, the fifth group of WR-TF24-OCSVM integrates wavelet reconstruction, raw signal statistical methods and OCSVM, and the sixth group of WR-CAE-OCSVM combines wavelet reconstruction, convolutional autoencoder and OCSVM.

4.3 Result analysis

4.3.1 Comparison with other models

Comparing multiple current anomaly detection models, We chose the variant autoencoder (VAE)-based anomaly detection method, which is currently widely used and influential in heart sound anomaly detection or related fields, is selected as the comparison object. The data used in the comparison experiments are from the same source as those used to validate the WCOS model in the dissertation, and the software and hardware environments run are the same to ensure the consistency and comparability of the data.

In the following, the heart sound detection error of the VAE model will be verified first, and the mean square error (MSE) will be chosen as the evaluation index, as shown in the following formula Equation 13:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

where n represents the number of samples, y_i represents the true value of the i th sample, and \hat{y}_i represents the predicted value of the i th sample.

The results of utilizing the five-fold method for the VAE model results are shown in Figure 4.

The VAE models for the three convolutional kernel cases can be seen in Figure 4. vae1 in Figure 4 represents the Test MSE for the VAE model with convolutional kernel of Wang et al. (2024) and Liang et al. (2021) is 1.257207727, vae2 represents the Test MSE for convolutional kernel of Liang et al. (2021) and Chong et al. (2021) is 1.259705973, and vae3 represents the Test MSE for convolutional kernel of Chong et al. (2021) and Mir et al. (2024) is 1.257824469. The smallest MSE is 1.257207727 when the number of convolutional kernels is (Wang et al., 2024; Liang et al., 2021). The maximum MSE is 1.259705973 when the number of convolution kernels is (Liang et al., 2021; Chong et al., 2021). The range of model MSE is during the range of [1.2572, 1.2597].

To compare with the WCOS model proposed in this paper, the root mean square error obtained by combining the VAE model with wavelet variations, a one-dimensional support vector machine, and using the five-fold method is shown below.

Figure 5 represents the MSE of WR-VAE-OCSVM for the six convolutional kernel cases, and the Test MSE of the VAE model with the convolutional kernel of Wang et al. (2024) and Liang et al. (2021) is 1.257626295, that of the convolutional kernel of Liang et al. (2021) and Chong et al. (2021) is 1.257779431, that of the convolutional kernel of Chong et al. (2021) and Mir et al. (2024) is 1.258005643, that of the convolutional kernel for Mir et al. (2024) and Li et al. (2024) has a Test MSE of 1.256320214, the Test MSE for convolution kernel for [32, 64] has a Test MSE of 1.256026459, and the Test MSE for convolution kernel for [64, 128] has a Test MSE of 1.259554124, and it can be seen that the model error is minimized at convolution kernel for [32, 64], when the Test MSE is 1.256026459, and when the convolution kernel is [64, 128], the model error is the largest, at this time the Test MSE is 1.259554124, and the error range is in [1.256026459, 1.259554124].

Figure 6 represents the MSE of WR-CAE-OCSVM for the six convolution kernel cases, and the Test MSE of the VAE model with convolution kernel (Wang et al., 2024; Liang et al., 2021) is 1.076472855, the Test MSE with convolution kernel (Liang et al., 2021; Chong et al., 2021) is 1.055950403, the Test MSE with convolution kernel (Chong et al., 2021; Mir et al., 2024) is 1.048900998, and the Test MSE of the VAE model with convolution kernel [16, 32] is

TABLE 1 Control group setup.

Control group no	Control model structure
1	OCSVM Model
2	TF24-OCSVM
3	CAE-OCSVM
4	WR-OCSVM
5	WR-TF24-OCSVM
6	WR-CAE-OCSVM

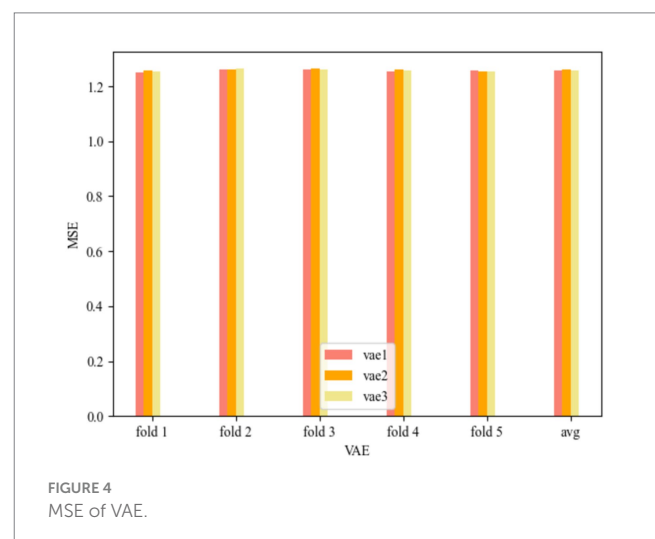


FIGURE 4 MSE of VAE.

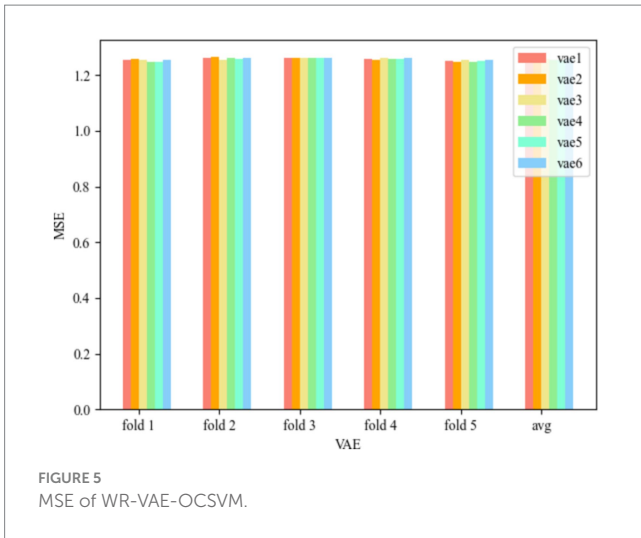


FIGURE 5 MSE of WR-VAE-OCSVM.

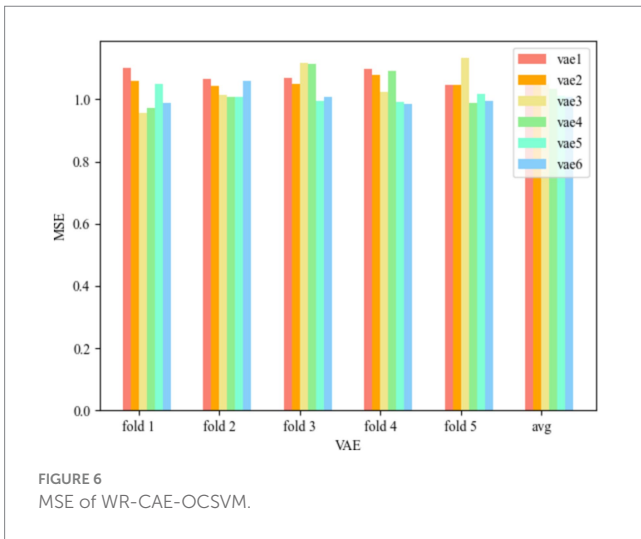


FIGURE 6 MSE of WR-CAE-OCSVM.

1.034770775, Test MSE for convolution kernel of [32, 64] is 1.012387657, and Test MSE for convolution kernel of [64, 128] is 1.007082856, which can be seen that the model error is minimum when the convolution kernel is [64, 128], and at this time Test MSE is 1.007082856, and the model error is maximum when the convolution kernel is (Wang et al., 2024; Liang et al., 2021), which is 1.007082856. At this time, the Test MSE is 1.076472855, and the error range is [1.007082856, 1.076472855].

Comparing the two models, it is obvious that the error of the model proposed in this paper is much lower than that of WR-VAE-OCSVM, and compared with the WR-VAE-OCSVM model, the minimum error is lower by 0.248944, which reduces it by 19.82%, and the accuracy of the WCOS model still performs better in a variety of test rounds, which verifies the validity and advancement of the model. To improve the detection accuracy of the model, the subsequent content will discuss the effects of hyperparameters in the model and find the optimal parameters, respectively.

4.3.2 Hyperparameter determination

The modules that need to determine hyperparameters include the OCSVM module and the convolutional autoencoder module. The

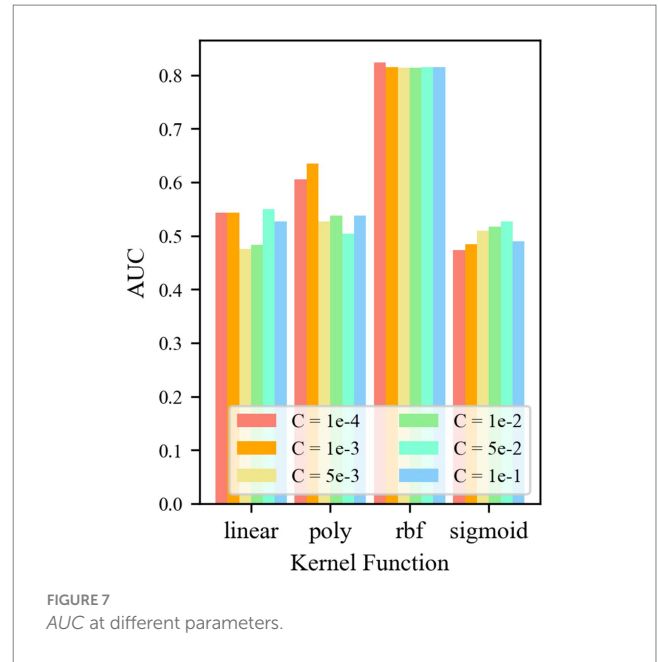


FIGURE 7 AUC at different parameters.

hyperparameters that the OCSVM module needs to determine are the control factor C and the kernel function k . Control factor Select $C = 1 \times 10^{-4}$, 1×10^{-3} , 5×10^{-3} , 1×10^{-2} , 5×10^{-2} and 1×10^{-1} performance. Commonly used kernel function is linear function, polynomial function, radial basis function, the sigmoid function, formula by $k(h_i, h_j) = \varphi(h_i)^T \varphi(h_j)$ is given by Equation 14:

$$\begin{aligned}
 k_{\text{line}}(h_i, h_j) &= h_i^T h_j \\
 k_{\text{poly}}(h_i, h_j) &= (a_1 h_i^T h_j + b_1)^d \\
 k_{\text{rbf}}(h_i, h_j) &= \exp\left(-\frac{h_i - h_j^2}{\sigma^2}\right) \\
 k_{\text{sigmoid}}(h_i, h_j) &= \tanh(a_2 h_i^T h_j + b_2)
 \end{aligned}
 \tag{14}$$

Where a_1 and b_1 are polynomial coefficients, σ are width parameters, and a_2 and b_2 are coefficients of the sigmoid function. What the convolutional self-coder module needs to determine is the convolutional kernel i and j . This article arranged the 6 kinds of combinations, respectively $(I = 2, j = 4)$, $(I = 4, j = 8)$, $(I = 8, j = 16)$, $(I = 16, j = 32)$, $(I = 32, j = 64)$, $(I = 64, j = 128)$. The results of each model are as follows.

For the control group OCSVM, Figure 7 shows that the AUC value of the OCSVM model is higher only when the kernel function $k = k_{\text{rbf}}$. When the kernel functions $k = k_{\text{linear}}$, $k = k_{\text{poly}}$, and k_{sigmoid} , the AUC values of the OCSVM model are lower. When the kernel function $k = k_{\text{rbf}}$, the AUC value is less affected by the control factor C . When $k = k_{\text{rbf}}$ and $C = 1 \times 10^{-4}$, the OCSVM model achieved the best performance in this task ($AUC = 0.855$).

For control group No. 2 TF24-OCSVM, Figure 8 shows that when kernel functions $k = k_{\text{rbf}}$ and k_{linear} , the AUC value of the TF24-OCSVM model is higher. When the kernel functions $k = k_{\text{poly}}$ and k_{sigmoid} , the AUC value of the TF24-OCSVM model is low. When the kernel function $k = k_{\text{rbf}}$, the AUC value is less affected by the control factor

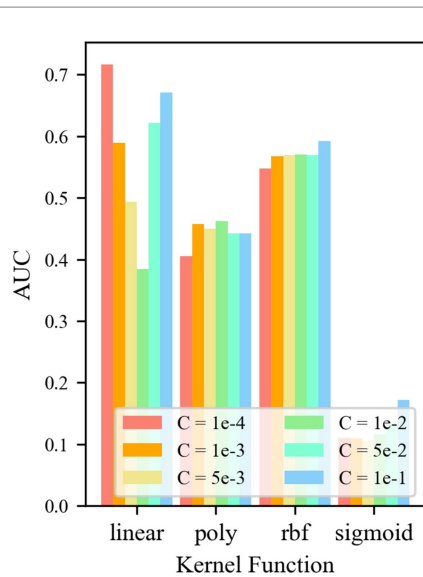


FIGURE 8
AUC for different parameters of TF24-OCSVM.

C. When the kernel function $k = k_{\text{linear}}$, the AUC value has a greater influence on the control factor C . When $k = k_{\text{linear}}$ and $C = 1 \times 10^{-4}$, the TF24-OCSVM model in this task achieved the best performance ($AUC = 0.716$).

For the control group No. 3 CAE-OCSVM, the results are shown in Figure 9. Figure 9 shows that when $k = k_{\text{rbf}}$, the AUC of the CAE-OCSVM model is significantly higher than the AUC of $k = k_{\text{linear}}$, k_{poly} , and k_{sigmoid} . Therefore, the kernel function of the CAE-OCSVM model is set to k_{rbf} . In this case, when the kernel is 2–4, the maximum AUC is 0.736 ($C = 1 \times 10^{-3}$). When the kernel is 4–8, the maximum AUC is 0.867 ($C = 1 \times 10^{-3}$). When the kernel is 8–16, the maximum AUC is 0.841 ($C = 1 \times 10^{-3}$). When the Kernel is 16–32, the maximum AUC is 0.842 ($C = 5 \times 10^{-2}$). When the kernel number is 32–64, the maximum AUC is 0.846 ($C = 1 \times 10^{-3}$). When the kernel is 86–128, the maximum AUC is 0.852 ($C = 1 \times 10^{-3}$). Therefore, to achieve the best performance of the CAE-OCSVM model in this task, the kernel function is set to $k = k_{\text{rbf}}$, the convolution kernel is set to 4–8, and the control factor is set to $C = 1 \times 10^{-3}$.

For control group No. 4 WR-OCSVM, the results are shown in Figure 10. Figure 10 shows that when kernel function $k = k_{\text{rbf}}$, the minimum AUC value of the WR-OCSVM model reaches 0.814. When the kernel function $k = k_{\text{linear}}$, the maximum AUC of the WR-OCSVM model reaches 0.538. When the kernel function $k = k_{\text{poly}}$, the maximum AUC value of the WR-OCSVM model reaches 0.601. When the kernel function $k = k_{\text{sigmoid}}$, the maximum AUC value of the WR-OCSVM model reaches 0.526. Therefore, when the kernel function $k = k_{\text{rbf}}$, the lowest AUC value is significantly higher than the highest AUC value when the kernel function $k = k_{\text{linear}}$, k_{poly} , k_{sigmoid} . When the kernel function $k = k_{\text{rbf}}$ and the control factor $C = 1 \times 10^{-4}$, the AUC value is the highest, reaching 0.824. Therefore, to achieve the best performance of the WR-OCSVM model in this task, the kernel function is set to $k = k_{\text{rbf}}$, and the control factor is set to $C = 1 \times 10^{-4}$.

For the control group No. 5 WR-TF24-OCSVM, the results are shown in Figure 11. It is shown that when the kernel functions $k = k_{\text{rbf}}$ and k_{linear} , the AUC value of the TF24-OCSVM model is higher. When the kernel functions $k = k_{\text{poly}}$ and k_{sigmoid} , the AUC value

of the TF24-OCSVM model is low. When the kernel function $k = k_{\text{rbf}}$, the AUC value is less affected by the control factor C . When the kernel function $k = k_{\text{linear}}$, the AUC value is greatly affected by the control factor C . When $k = k_{\text{linear}}$ and $C = 1 \times 10^{-4}$, the TF24-OCSVM model in this task achieved the best performance ($AUC = 0.796$).

For the control group No. 6 WR-CAE-OCSVM, the results are shown in Figure 12: when $k = k_{\text{rbf}}$, the AUC of the WR-CAE-OCSVM model is significantly higher than that when $k = k_{\text{linear}}$, k_{poly} , and k_{sigmoid} . Therefore, the kernel function of the WR-CAE-OCSVM model is set to k_{rbf} . In this case, when the kernel is 2–4, the maximum AUC is 0.779 ($C = 1 \times 10^{-3}$). When the kernel is 4–8, the maximum AUC is 0.822 ($C = 1 \times 10^{-4}$). When the kernel is 8–16, the maximum AUC is 0.848 ($C = 1 \times 10^{-3}$). When the kernel is 16–32, the maximum AUC is 0.850 ($C = 1 \times 10^{-2}$). When the kernel is 32–64, the maximum AUC is 0.851 ($C = 1 \times 10^3$). When the kernel is 86–128, the maximum AUC is 0.870 ($C = 1 \times 10^{-3}$). Therefore, to achieve the best performance of the WRCAE-OCSVM model in this task, the kernel function is set to $k = k_{\text{rbf}}$, the convolutional kernel is set to 64–128, and the control factor is set to $C = 1 \times 10^{-3}$.

In summary, the optimal hyperparameter Settings of the six groups of models are shown in Table 2.

The table comprehensively shows the optimal hyperparameter settings of the six groups of models. The models are carefully tuned for different model structural characteristics, such as the OCSVM model kernel function is set to $k = k_{\text{rbf}}$ and the control factor $C = 0.0001$ in control group 1, the kernel function is set to $k = k_{\text{linear}}$ and the control factor $C = 0.0001$ in control group 2, and the convolutional kernel is set to Liang et al. (2021) and Chong et al. (2021) for the CAE-OCSVM model in control group 3, the kernel function is set to Liang et al. (2021) and Chong et al. (2021), and the control factor $C = 0.0001$ in control group 4, the kernel function is set to $k = k_{\text{rbf}}$ control factor $C = 0.0001$, as in control group 4 WR-OCSVM model kernel function is set to $k = k_{\text{rbf}}$ control factor $C = 0.0001$, as in control group 5 WR-TF24-OCSVM model kernel function is set to $k = k_{\text{linear}}$, control factor $C = 0.0001$, as in control group 6 WR-CAE-OCSVM model convolution kernel is set to [64, 128], the kernel function is set to $k = k_{\text{rbf}}$ control factor $C = 0.0001$.

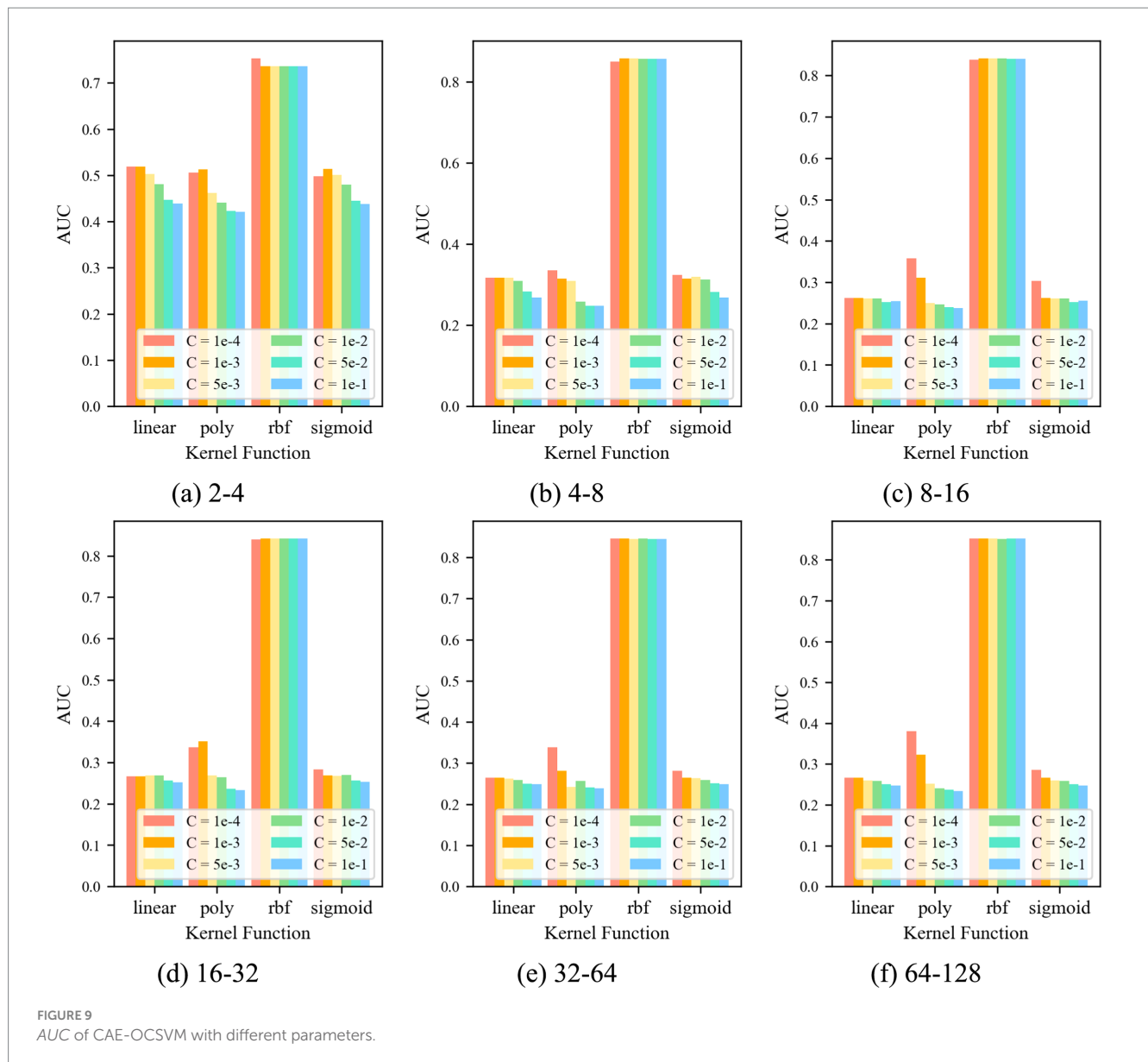
4.3.3 Effectiveness of the developed framework

This paper established five groups of comparison experiments based on the cross-difference verification method to verify the validity. To remove the randomness of the neural network, each experiment was repeated 10 times. The experimental results are shown in Figure 13.

Figure 13 compares the AUC values of samples of six models OCSVM, TF24-OCSVM, CAE-OCSVM, WR-OCSVM, WR-TF24OCSVM, and WR-CAE-OCSVM under the best hyperparameters. It shows that the WR-CAE OCSVM has the highest AUC value in the samples with $K = 1, 2, 3$, and 4, reaching 0.867, 0.878, 0.886, 0.842, and 0.875, respectively.

Table 3 shows the Statistics of the mean, standard deviation, and variance of AUC of OCSVM, TF24-OCSVM, CAE-OCSVM, WR-OCSVM, WR-TF24-OCSVM and WR-CAE-OCSVM models, respectively.

Table 3 shows that the average AUC of WR-CAE-OCSVM is higher than that of WR-TF24-OCSVM (0.870–0.796/



$0.870 \times 100\% = 8\%$. Similarly, the average AUC of WR-CAE-OCSVM is 5, 1, 18, and 5% higher than that of WR-OCSVM, CAE-OCSVM, TF24-OCSVM, and OCSVM, respectively. The above data reflect that WR-CAE-OCSVM has the best effect in this heart sound classification task.

Table 3 also shows that the AUC standard deviation of WR-CAE-OCSVM is lower than that of WR-TF24-OCSVM ($(0.0938-0.0169)/0.0938 \times 100\% = 82\%$). Similarly, the AUC standard deviation of WR-CAE-OCSVM is 42, 55, 87, and 42% lower than that of WROCSVM, CAE-OCSVM, TF24-OCSVM, and OCSVM, respectively. The AUC variance of WR-CAE-OCSVM is 97, 66, 80, 98, and 66% lower than that of WR-TF24-OCSVM, WR-OCSVM, CAE-OCSVM, TF24-OCSVM, and OCSVM, respectively. The above data reflect that the WR-CAE-OCSVM model has the best stability in this heart sound classification task.

4.3.4 Anti-noise ability of the developed framework

To verify the anti-noise ability of the WR-CAE-OCSVM model in the classification of heart sounds, four groups of experiments were carried out, respectively. In the experiment, Gaussian noise with different standard variance (σ) was added to heart sounds to simulate ambient noise. As shown in Figure 14, gaussian noise with four different σ values is set.

Five groups of experiments were established based on the cross-difference verification method, and four groups of noise were added to each group of experimental data. To eliminate the randomness of the neural network, each experiment was repeated 10 times, and the experimental results are shown in Figure 15.

Figure 15 shows that the WR-CAE-OCSVM has the best classification effect in this task. When the noise of $\sigma = 0$, the mean AUC of WR-CAE-OCSVM is higher than that of

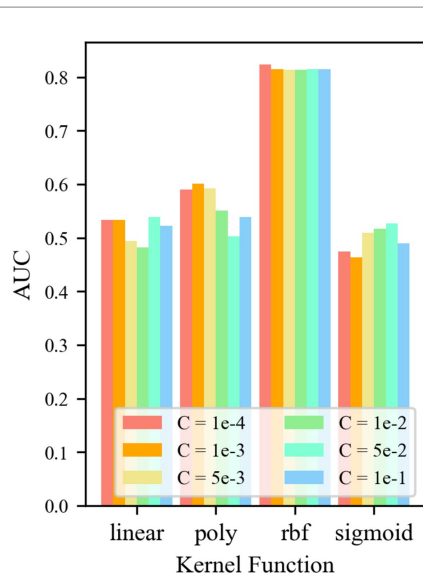


FIGURE 10
AUC at different parameters.

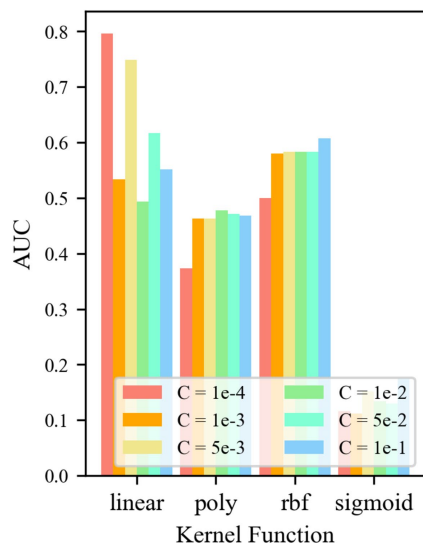


FIGURE 11
AUC at different parameters.

WR-OCSVM (0.870–0.824)/ 0.870 × 100% = 5.3%. The mean AUC of WR-CAE-OCSVM was 1.4 and 5.3% higher than that of CAE-OCSVM and OCSVM, respectively. When noise with sigma = 0.1, the mean AUC of the WR-CAE-OCSVM is 4.6, 0.2, and 7.2% higher than that of WR-OCSVM, CAEOCSVM and OCSVM, respectively. When noise with sigma = 0.25, the mean AUC of the WR-CAE-OCSVM is 7.4, 14.7, and 20.3% higher than that of WR-OCSVM, CAE-OCSVM and OCSVM, respectively. When noise with sigma = 0.5, the mean AUC of the WR-CAE-OCSVM is 5.0, 6.6, and 16.5% higher than that of WR-OCSVM, CAE-OCSVM, and OCSVM, respectively. To verify the stability of each model, Figure 16 shows the standard deviation of the AUC obtained from five sets of experiments.

Figure 16 shows that the WR-CAE-OCSVM model has the best stability in this task. When the noise of sigma = 0, the AUC standard deviation of WR-CAE-OCSVM is lower than the mean AUC of the WR-OCSVM (0.0290–0.0169)/0.0290 × 100% = 41.8%. The AUC standard deviation of the WR-CAE-OCSVM is 55.4 and 41.8% lower than that of CAE-OCSVM, and OCSVM, respectively. When the noise of sigma = 0.1, the AUC standard deviation of the WR-CAE-OCSVM is 44.9, 34.3, and 57.2% lower than that of WR-OCSVM, CAE-OCSVM, and OCSVM, respectively. When the noise of sigma = 0.25, the AUC standard deviation of the WR-CAE-OCSVM is 20.7, 58.9, and 19.8% lower than that of WR-OCSVM, CAE-OCSVM, and OCSVM, respectively. When the noise of sigma = 0.5, the AUC standard deviation of the WR-CAE-OCSVM is 19.2, 54.1, and 29.8% lower than that of WR-OCSVM, CAE-OCSVM and OCSVM, respectively.

5 Discussion and future perspectives

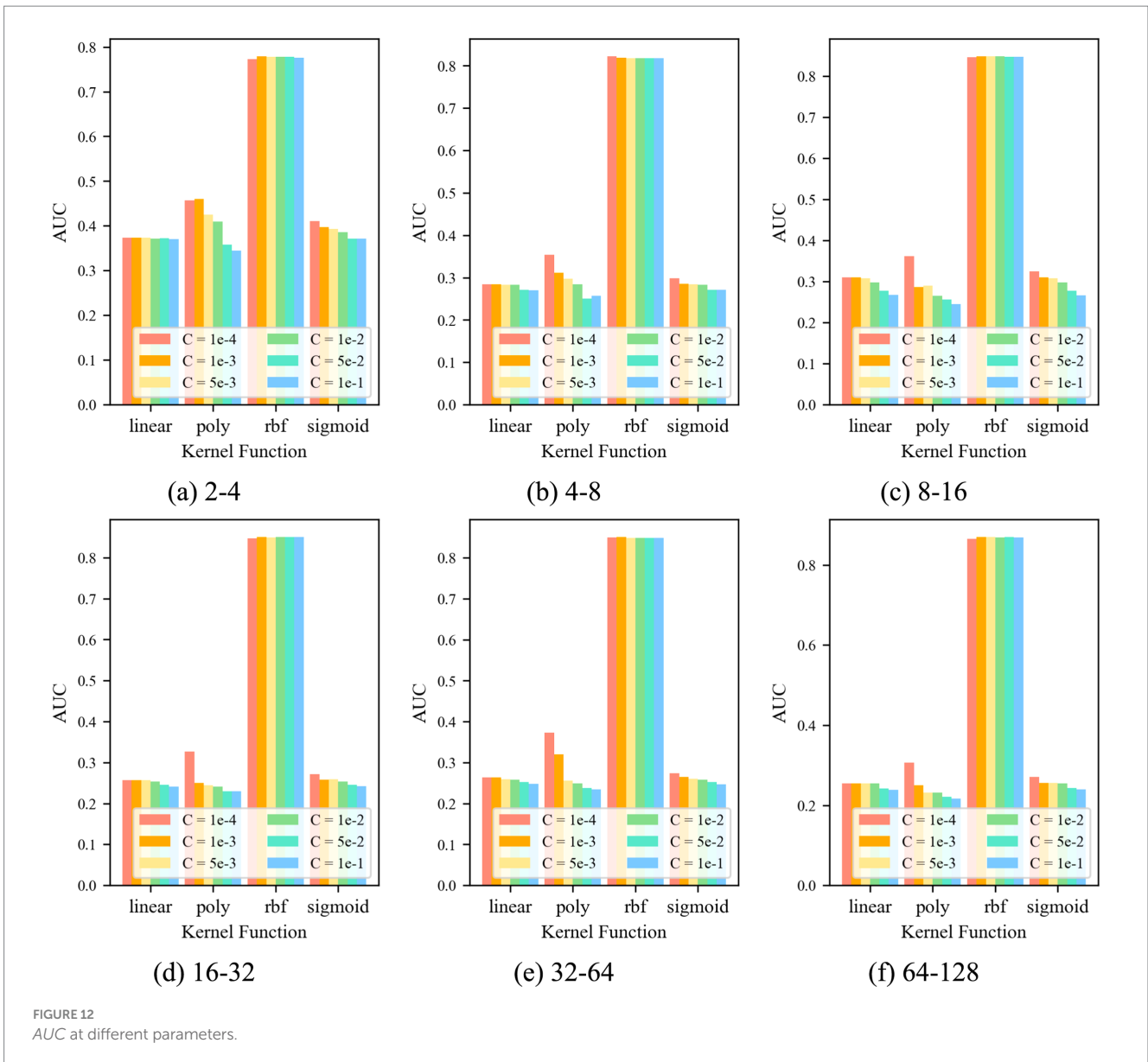
This paper presents a new semi-supervised anomaly detection method WCOS, which can work in both unsupervised and semi-supervised Settings. Experimental evaluations using real data sets show that WCOS is much more accurate than current anomaly detection methods and has greater noise resistance after adding different noise effects. In future work, we will improve the clustering process and anomaly measurement to make the detection more efficient and time-saving. Surveillance data is often uneven, so our future efforts will focus on discovering a suitable regularization technique that may further improve the stability of the anomaly detection model.

5.1 Advantages of the WCOS framework

In this paper, we propose a new semi-supervised anomaly detection method, WCOS, that can work in both unsupervised and semi-supervised settings. It has several significant advantages over other anomaly detection methods, firstly its semi-supervised nature is a key advantage, in the real world, obtaining a large amount of labeled data for heart sound anomaly detection is often both arduous and expensive, WCOS can effectively utilize both limited labeled data and a large amount of unlabeled data, by training a convolutional autoencoder with normal samples and using a one-class support vector machine, it can learn the normal heart sound's inherent patterns and accurately differentiate abnormal heart sounds in the

TABLE 2 Hyperparameter setting.

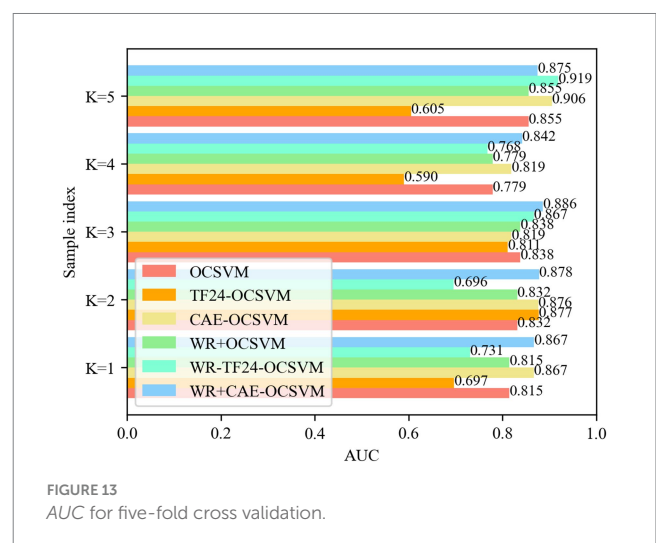
Control group number	Control group model structure	Hyperparameter setting
1	OCSVM Model	OCSVM (rbf-0.0001)
2	TF24-OCSVM	OCSVM (linear-0.0001)
3	CAE- OCSVM	CAE-OCSVM (4-8-rbf-0.001)
4	WR-OCSVM	OCSVM (rbf-0.0001)
5	WR-TF24-OCSVM	OCSVM (linear-0.0001)
6	WR- CAE -OCSVM	CAE-OCSVM (64-128-rbf-0.001)



presence of sparse labeled abnormal samples, which greatly improves its practical applicability in the clinical setting.

Secondly, the integration of wavelet reconstruction provides significant noise suppression capability, heart sound signals are often contaminated by various noises, such as power line interference and baseline drift, wavelet reconstruction decomposes the signal into approximation coefficients and detail coefficients, and by eliminating the detail coefficients that mainly contain noise components, it effectively filters out the high-frequency noises while preserving the basic low-frequency trends and characteristics of heart sounds, and this denoised signal can be used as a subsequent This denoised signal can be used as a high-quality input for subsequent analysis, improving signal quality and the reliability of detection results.

The efficacy of the framework is further enhanced by the combination of a convolutional autoencoder and one classification support vector machine. The convolutional autoencoder has a convolutional and deconvolutional layer that extracts hierarchical abstract features from the heart sound data, which not only reduces



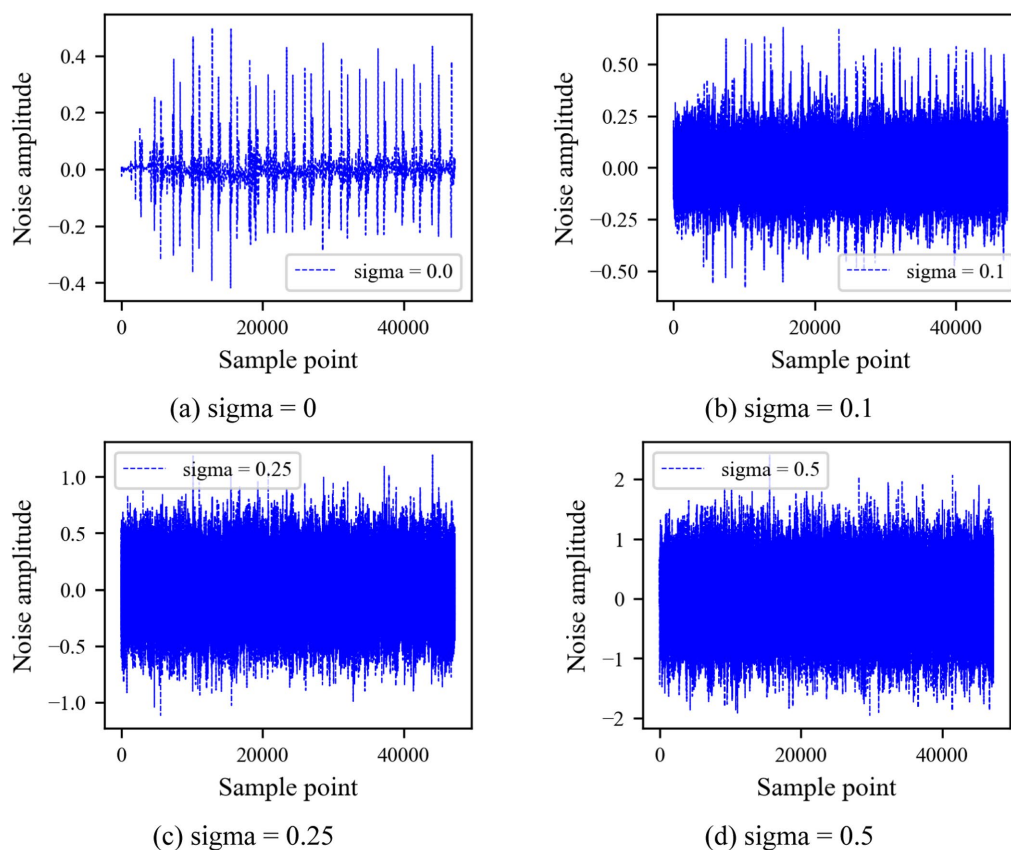


FIGURE 14
Four Gaussian noises with different sigma values.

the dimensionality of the data but also captures underlying patterns and correlations in the signal. One classification support vector machine then utilizes these learned features to establish accurate decision boundaries to distinguish between normal and abnormal heart sounds, and experimental results on real datasets fully demonstrate the effectiveness of WCOS.

5.2 Disadvantages and limitations

Despite its merits, the WCOS framework has certain drawbacks. The computational complexity during the training process is relatively high. The training of the convolutional autoencoder demands significant computational resources and time, particularly when handling large-scale heart sound datasets. The determination of optimal hyperparameters for the one-class support vector machine also requires extensive experimentation and fine-tuning, adding to the computational burden.

Moreover, while the model performs well in the current study focused on heart sound anomaly detection, its generalization to other types of medical signals or datasets with different characteristics may be limited. The model is trained and optimized based on the specific features and distributions of the heart sound data used in this research. When applied to other medical signal domains, such as electroencephalogram (EEG) or electromyogram (EMG) signals, or datasets with distinct statistical properties, its performance may

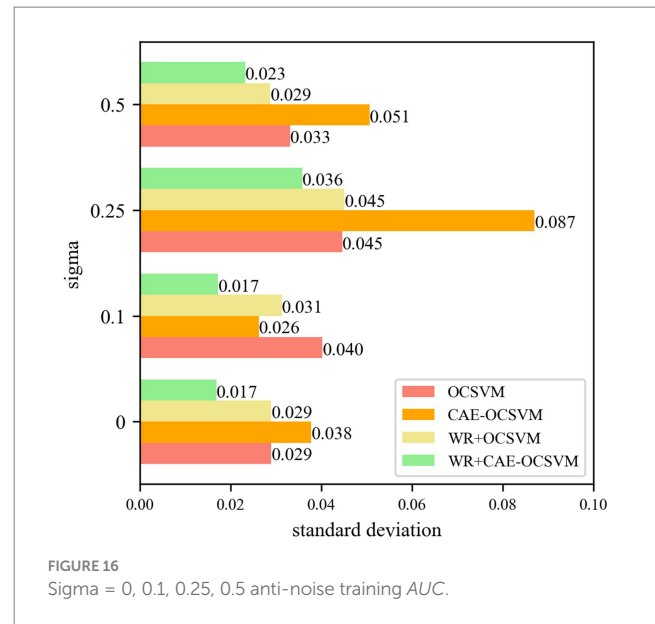
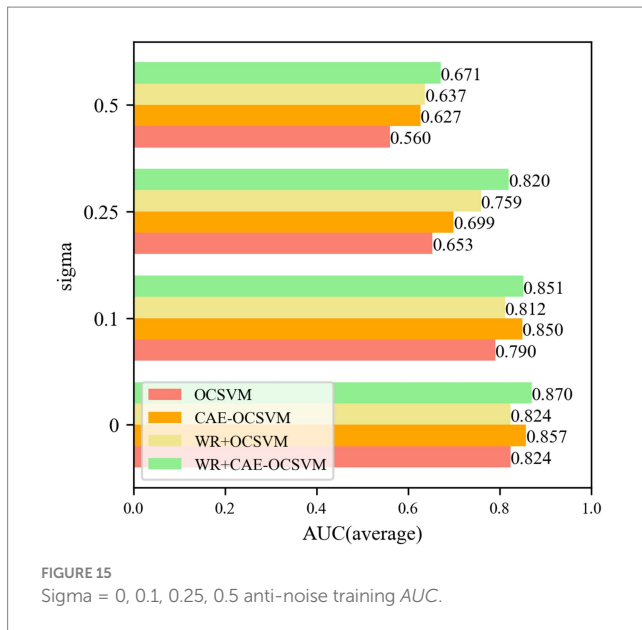
TABLE 3 Mean, standard deviation, and variance of AUC for 6 models.

Model	AUC mean	AUC standard deviation	AUC variance
OCSVM	0.824	0.0290	0.0008
TF24-OCSVM	0.716	0.1260	0.0159
CAE-OCSVM	0.857	0.0378	0.0014
WR-OCSVM	0.824	0.0290	0.0008
WR-TF24-OCSVM	0.796	0.0938	0.0088
WR-CAE-OCSVM	0.870	0.0169	0.0003

decline due to differences in signal characteristics, noise patterns, and data distributions.

5.3 Future research directions

To address these limitations, future research efforts will be directed toward several aspects. In terms of computational efficiency, we plan to explore advanced model compression techniques. These techniques could involve pruning the redundant connections and neurons in the convolutional autoencoder without



significantly sacrificing its performance. Additionally, we will investigate the application of parallel computing architectures to accelerate the training process. By distributing the computational tasks across multiple processors or computing devices, we can reduce the training time and make the framework more accessible for real-time or large-scale applications.

Regarding generalization, we intend to conduct more extensive cross-dataset validations. This will involve collaborating with other research institutions to access diverse medical signal datasets from different sources and populations. By evaluating the WCOS framework on these varied datasets, we can identify its strengths and weaknesses in different contexts and make necessary adjustments. Furthermore, we will explore the incorporation of domain adaptation algorithms. These algorithms can help the model adapt to new data domains by learning the transferable features and reducing the domain shift between the training and target datasets. This will enhance the model's ability to generalize to a broader range of medical signal analysis tasks and contribute to the development of more robust and versatile anomaly detection methods in the medical field.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: data is unavailable due to privacy. Requests to access these datasets should be directed to Shuimiao Kang, 2022095018@cauc.edu.cn.

Author contributions

PZ: Conceptualization, Data curation, Formal analysis, Investigation, Writing – review & editing. SK: Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. FF: Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft,

Writing – review & editing. JL: Investigation, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Civil Aviation Safety Capacity Building Fund (No. [2024]28), BCH Young Investigator Program Fund (No. 3-1-014-01-13), and Central University Project Fund (No. KJZ53420230039).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aziz, S., Khan, U. M., Iqtidar, K., and Fernandez-Rojas, R. (2024). Diagnosis of schizophrenia using EEG sensor data: a novel approach with automated log energy-based empirical wavelet reconstruction and cepstral features. *Sensors* 24:6508. doi: 10.3390/s24206508
- Chen, S., Sun, Y., Wang, J., Wan, M., Liu, M., and Li, X. (2024). A multi-scale dual-decoder autoencoder model for domain-shift machine sound anomaly detection. *Digit. Signal Process.* 104813.
- Chong, Z., Pang, S., and Zheng, W. S. (2021). A survey on deep learning-based anomaly detection. *ACM Comput. Surveys* 54, 1–41.
- Cloudera, and Nisha, M. (2020). "Deep Learning for Anomaly Detection". Report, Cloudera.
- Fan, H., Ye, J., Chen, W., and Chen, Z. (2021). One-class variational autoencoder for anomaly detection in text data. *IEEE Access* 9, 11483–11492.
- Gangloff, H., Pham, T. M., Courtrai, L., and Lefèvre, S. (2024). Variational autoencoder with Gaussian random field prior: application to unsupervised animal detection in aerial images. *ISPRS J. Photogramm. Remote Sens.* 218, 600–609.
- Gao, P., Xu, Q., and Zhang, C. (2021). Anomaly detection in multivariate time series with conditional generative flow. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 3257–3268.
- Jain, P., Tiwari, A. K., and Som, T. (2022). An intuitionistic fuzzy bireduct model and its application to cancer treatment. *Comput. Ind. Eng.* 168:108124. doi: 10.1016/j.cie.2022.108124
- Jain, P., Tiwari, A., and Som, T. (2023). Fuzzy rough assisted missing value imputation and feature selection. *Neural Comput. & Applic.* 35, 2773–2793. doi: 10.1007/s00521-022-07754-9
- Jia-Yan, G., Rong-Hua, L., Yan, Z., and Guo-Ren, W. (2020). A dynamic network anomaly detection algorithm based on graph neural networks. *J. Softw.* 31, 748–762.
- Jyothi, P., and Pradeepini, G. (2021). Classification of normal/abnormal heart sound recording through convolution neural network through the integration of baseline and adaboost classifier. In Proceedings of the 2nd International Conference on Computational and Bio Engineering: CBE 2020. Singapore: Springer Singapore. 1395–1414.
- Kumar, R., Singh, U. P., Bali, A., Chouhan, S. S., and Tiwari, A. K. (2024). Adaptive control of unknown fuzzy disturbance-based uncertain nonlinear systems: application to hypersonic flight dynamics. *J. Analysis* 32, 1395–1414. doi: 10.1007/s41478-023-00687-z
- Li, Q. (2024). Electricity behavior modeling and anomaly detection services based on a deep variational autoencoder network. *Energies* 17. doi: 10.3390/en17163904
- Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., et al. (2023). A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection. *IEEE Trans. Inf. Forensics Secur.* 18, 1638–1652. doi: 10.1109/TIFS.2023.3245413
- Li, X., Zhou, W., Yin, J., Zhang, Z., Huang, G., Sheng, Y., et al. (2024). Convolutional variational autoencoder and multi-scale attention convolutional neural network based diagnostics on filament current sensors for mass spectrometers. *Eng. Appl. Artif. Intell.* 138:109443. doi: 10.1016/j.engappai.2024.109443
- Liang, X. W. R. D., and F. H. Y. (2021). Unsupervised deep anomaly detection model based on sample association perception. *J. Comput.* 44, 2317–2331.
- Ling, C. X., Huang, J., and Zhang, H. (2003). AUC: a statistically consistent and more discriminating measure than accuracy. *In: Jicai*, 3, 519–524.
- Liu, S., Qin, Z., Gan, X., and Wang, Z. (2019). "SCOD: a novel semi-supervised outlier detection framework" in 2019 IEEE/CIC international conference on Communications in China (ICCC) (China: Changchun), 316–321.
- Mir, A. A., Zuhairi, F. M., and Musa, S. M. (2023). Graph anomaly detection with graph convolutional networks. *Int. J. Advan. Comput. Sci. Applic.* 14.
- Mir, A. A., Zuhairi, M. F., Musa, S., and Namoun, A. (2024). Adaptive anomaly detection in dynamic graph networks. In Adaptive Anomaly Detection in Dynamic Graph Networks, IEEE. 200–206.
- Niu, M., Wang, Y., Song, K., Wang, Q., Zhao, Y., and Yan, Y. (2021). An adaptive pyramid graph and variation residual-based anomaly detection network for rail surface defects. *IEEE Trans. Instrum. Meas.* 70, 1–13. doi: 10.1109/TIM.2021.3125987
- Rezaee, K., Khosravi, M. R., Jabari, M., Hesari, S., Anari, M., and Aghaei, F. (2022). Graph convolutional network-based deep feature learning for cardiovascular disease recognition from heart sound signals. *Int. J. Intell. Syst.* 37, 11250–11274. doi: 10.1002/int.23041
- Ruff, L., Vandermeulen, R. A., Goernitz, N., Deecke, L., and Kloft, M. (2021). Deep one-class classification. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2298–2308.
- Sanders, K., Danielczuk, M., Mahler, J., Tanwani, A., and Goldberg, K. (2020). Non-Markov policies to reduce sequential failures in robot bin picking." 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Hong Kong, China, 2020, pp. 1141–1148.
- Shen, Y., Wang, J., and Gao, J. (2021). Noise suppression for vector magnetic anomaly detection by noise spatial characteristics investigation. *IEEE geoscience and remote sensing letters*, 19, 1–4.
- Shi, Y., Hou, B., Liu, J., Liu, A., Guo, S., and Liu, J. (2023). Element defective sample augmentation method based on improved DCGAN," 2023 IEEE 16th International Conference on Electronic Measurement & Instruments (ICEMI), Harbin, China, pp. 167–171.
- Shreevastava, S., Maratha, P., Som, T., and Tiwari, A. K. (2023). A novel (alpha, beta)-indiscernibility-assisted intuitionistic fuzzy-rough set model and its application to dimensionality reduction. *Optimization*, 1–21.
- Wang, Y., Yang, X., Ye, M., Zhao, Y., Chen, R., Da, M., et al. (2024). Machine learning-based intelligent auscultation techniques in congenital heart disease: application and development. *Congenital Heart Disease*. 19.
- Xiao-dong, N., Li-Rong, L., and Ji-Rong, W. (2020). Reconstruction of ECG signals based on improved empirical modal decomposition domain inner physical feature recognition mode components. *J. Phys.* 70, 311–319.
- Xing, W., Kun, W., and Jian, W. (2020). Research on anomaly detection method based on deep generative model. Tianjin: Civil Aviation University of China.
- Xue, Q., Kun, W., and Ji-Rong, Q. (2022). Research progress on the development mechanism of malnutrition in children with congenital heart disease. *Electr. J. Modern Med. Health Res.* 6:4.
- Yuan, X., Ning, K., Wei, K., He, Y., Zhu, Q. X., et al. (2024). A hybrid MTS anomaly detection method based on reconstruction and adaptive spatial-temporal graph network. In 2024 IEEE 13th data driven control and learning systems conference (DDCLS). IEEE. 263–267.
- Zhang, X. Y., Zhang, X. P., Yu, H. G., and Liu, Q. S. (2025). A confident learning-based support vector machine for robust ground classification in noisy label environments. *Tunnel. Underground Space Technol. Incorporat. Trenchless Technol. Res.* 155:106128.