frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 08 January 2025
DOI 10.3389/frai.2024.1467218

Check for updates

OPEN ACCESS

EDITED BY
Dongpo Xu,
Northeast Normal University, China

REVIEWED BY
Sandeep Kumar Mishra,
Yale University, United States
Aili Wang,
Harbin University of Science and Technology, China
Junxiang Huang,
Boston College, United States

*CORRESPONDENCE
Suyu Dong
✉ dongsuyu@nefu.edu.cn

RECEIVED 24 July 2024
ACCEPTED 25 November 2024
PUBLISHED 08 January 2025

CITATION
Ma S, Zhang Y, Li D, Sun Y, Qiu Z, Wei L and Dong S (2025) SPEMix: a lightweight method via superclass pseudo-label and efficient mixup for echocardiogram view classification. *Front. Artif. Intell.* 7:1467218. doi: 10.3389/frai.2024.1467218

COPYRIGHT
© 2025 Ma, Zhang, Li, Sun, Qiu, Wei and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

SPEMix: a lightweight method via superclass pseudo-label and efficient mixup for echocardiogram view classification

Shizhou Ma[1], Yifeng Zhang[2], Delong Li[2], Yixin Sun[3], Zhaowen Qiu[2], Lei Wei[4] and Suyu Dong[2]*

[1]College of Aulin, Northeast Forestry University, Harbin, China, [2]College of Computer and Control Engineering, Northeast Forestry University, Harbin, China, [3]The Department of Ultrasound, Harbin Medical University Cancer Hospital, Harbin, China, [4]Department of Cardiovascular Surgery, First Affiliated Hospital With Nanjing Medical University, Nanjing, China

Introduction: In clinical, the echocardiogram is the most widely used for diagnosing heart diseases. Different heart diseases are diagnosed based on different views of the echocardiogram images, so efficient echocardiogram view classification can help cardiologists diagnose heart disease rapidly. Echocardiogram view classification is mainly divided into supervised and semi-supervised methods. The supervised echocardiogram view classification methods have worse generalization performance due to the difficulty of labeling echocardiographic images, while the semi-supervised echocardiogram view classification can achieve acceptable results via a little labeled data. However, the current semi-supervised echocardiogram view classification faces challenges of declining accuracy due to out-of-distribution data and is constrained by complex model structures in clinical application.

Methods: To deal with the above challenges, we proposed a novel open-set semi-supervised method for echocardiogram view classification, SPEMix, which can improve performance and generalization by leveraging out-of-distribution unlabeled data. Our SPEMix consists of two core blocks, DAMix Block and SP Block. DAMix Block can generate a mixed mask that focuses on the valuable regions of echocardiograms at the pixel level to generate high-quality augmented echocardiograms for unlabeled data, improving classification accuracy. SP Block can generate a superclass pseudo-label of unlabeled data from the perspective of the superclass probability distribution, improving the classification generalization by leveraging the superclass pseudolabel.

Results: We also evaluate the generalization of our method on the Unity dataset and the CAMUS dataset. The lightweight model trained with SPEMix can achieve the best classification performance on the publicly available TMED2 dataset.

Discussion: For the first time, we applied the lightweight model to the echocardiogram view classification, which can solve the limits of the clinical application due to the complex model architecture and help cardiologists diagnose heart diseases more efficiently.

KEYWORDS

superclass pseudo-label, lightweight, semi-supervised, open-set, echocardiogram view classification

1 Introduction

When diagnosing heart diseases based on the echocardiogram, different heart diseases depend on different views of the echocardiogram. For example, aortic stenosis can be diagnosed by analyzing the PLAX and PSAX views (Huang et al., 2021), and early myocardial infarction can be detected via A4C and A2C (Degerli et al., 2024). Hence, cardiologists usually need to identify the critical echocardiogram view during the clinical diagnostic process. Automated echocardiogram view classification can effectively reduce the clinical diagnosis time (Zhu et al., 2022). Figure 1 illustrates the clinical application process of the automated echocardiogram view classification. In practical clinical applications, automated view classification involved four steps. First, the cardiologist collected the echocardiogram data from the patient. Then, the cardiologist inputted the data into the automated view classification model to perform the echocardiogram view classification. Next, the cardiologist got the classification results of the echocardiogram. Finally, the cardiologist selected the expected view when making a diagnosis of a different disease and performed the diagnosis. Obviously, the core of automated echocardiogram view classification methods is to classify by the classification model. In this situation, developing a classification model that can identify the critical view precisely and efficiently will be vital in medicine. So many studies aim to develop echocardiogram view classification methods to assist cardiologists in diagnosing heart disease.

With the development of deep learning in medical images, there are many studies (Kusunose et al., 2020; Madani et al., 2018; Gao et al., 2017) focused on the classification of echocardiogram views by deep learning. These methods can be divided into supervised methods and semi-supervised methods.

Supervised studies used a large number of labeled data to train a model. Madani et al. (2018) performed the classification task by building a convolution model consisting of traditional convolution layers and fully connected layers. They classified 15 different echocardiogram views and got impressive classification results on their private dataset. Kusunose et al. (2020) collected 17,000 echocardiogram images from 340 patients and built a neural network model. The model consisted of five convolution layers and five pooling layers to classify them into five categories. Gao et al. (2017) fused the spatial and temporal information to perform eight view classifications based on their 432 video data. Although these supervised methods have gotten good results in identifying or detecting tasks (Gao et al., 2017; Han et al., 2024; Han et al., 2024), they need massive labeled data. Unfortunately, it will take a lot of time and energy for medical experts to label the medical images. Furthermore, these studies (Kusunose et al., 2020; Madani et al., 2018; Gao et al., 2017) used proprietary datasets, which made the generalization of their methods cannot be guaranteed. The private datasets also made it challenging to apply actual clinical needs. To address these issues, our study focused on the development of a semi-supervised algorithm within the publicly available dataset. We evaluated the generalization of our method in two different public datasets to prove the feasibility of our SPEMix in clinical application.

Semi-supervised studies (Madani et al., 2018; Hagberg et al., 2022) tried to decrease the cost of labeling data and improve efficiency by utilizing small labeled data and massive unlabeled data. Therefore, semi-supervised learning has been revealed to have superior potential (Xiaojin, 2008), and it has also been widely used in medicine (Chebli et al., 2018; Bai et al., 2017). Many scholars (Madani et al., 2018; Hagberg et al., 2022; Huang et al., 2023; Huang et al., 2024) have used semi-supervised learning for the echocardiogram view classification task. Madani et al. (2018) developed a semi-supervised generative adversarial network model to classify 15 views of echocardiogram via only a little labeled data. This method revealed the huge potential of semi-supervised. Hagberg et al. (2022) innovatively developed a semi-supervised learning method using natural language processing for right ventricle view classification, which improved the efficiency of classification. Although these methods had good classification efficiency, their classification accuracy was usually worse than the classification accuracy of supervised methods due to ignoring the damage caused by the out-of-distribution data. Most current methods have not achieved accurate and reliable classification accuracy to meet practical clinical needs. We proposed a novel efficient mixup method for our semi-supervised method to solve the above problem. Different from other mixing methods (Zhang et al., 2018; Yun et al., 2019; Liu et al., 2022) applied to natural images, our proposed efficient mixup method can automatically focus on the valuable pixel of the echocardiogram based on dynamic attention and produce high-quality augmented echocardiogram images. Specifically, we proposed a DAMix Block for our SPEMix to achieve our efficient augmentation.

Otherwise, many studies have observed that out-of-distribution of unlabeled data can harm the accuracy of semi-supervised learning (Oliver et al., 2018; Li et al., 2023; Zhao et al., 2022). Out-of-distribution data means the data does not belong to any of the known classification categories. The semi-supervised methods achieved good results only when the unlabeled data shared the same class space with labeled data. Unfortunately, the collected unlabeled medical image datasets often included out-of-distribution data in practical medical applications. This phenomenon usually leads to unlabeled data sharing a different class space with labeled data. Some approaches tried to improve the medical image classification accuracy via adversarial attacks and defenses (Lee et al., 2024; Kwon, 2023; Kwon and Lee, 2023). However, the dominant approaches to tackle the harm caused by the out-of-distribution data in natural images were to detect and filter the out-of-distribution data (Saito et al., 2021; Calderon-Ramirez et al., 2022). Huang et al. (2023) used augmentation and step direction modification to reduce the harm of out-of-distribution for echocardiogram view classification. The proposed method utilized the gradient information from unlabeled data from out-of-distribution only when the out-of-distribution data could improve the model performance. Different from the previous techniques, filtering out the out-of-distribution or modifying the gradient descent updates, we designed a novel, effective open-set framework to make better use of the out-of-distribution unlabeled data. Inspired by the success of the superclass work (Li et al., 2023) in solving the harm of out-of-distribution, we modeled the unlabeled data from superclass distribution. We assigned the superclass pseudo-label for unlabeled data. Specifically, we regarded all of the classes out of the distribution as a new superclass, and we designed the SP Block to generate superclass pseudo-labels to leverage the information in unlabeled datasets better. The SP Block included a multiclass classifier and a close-set classifier to calculate the out-of-distribution class probability distribution and the in-the-distribution class probability distribution. SP Block can calculate the superclass probability distribution and generate the superclass pseudo-labels based on these probabilities. Then, our superclass pseudo-labels can be leveraged by an open-set classifier, and our model can learn the semantic features of unlabeled data that are out of the distribution.

On the other hand, some researcher (Avola et al., 2024) focused on developing more complex models to improve the accuracy of

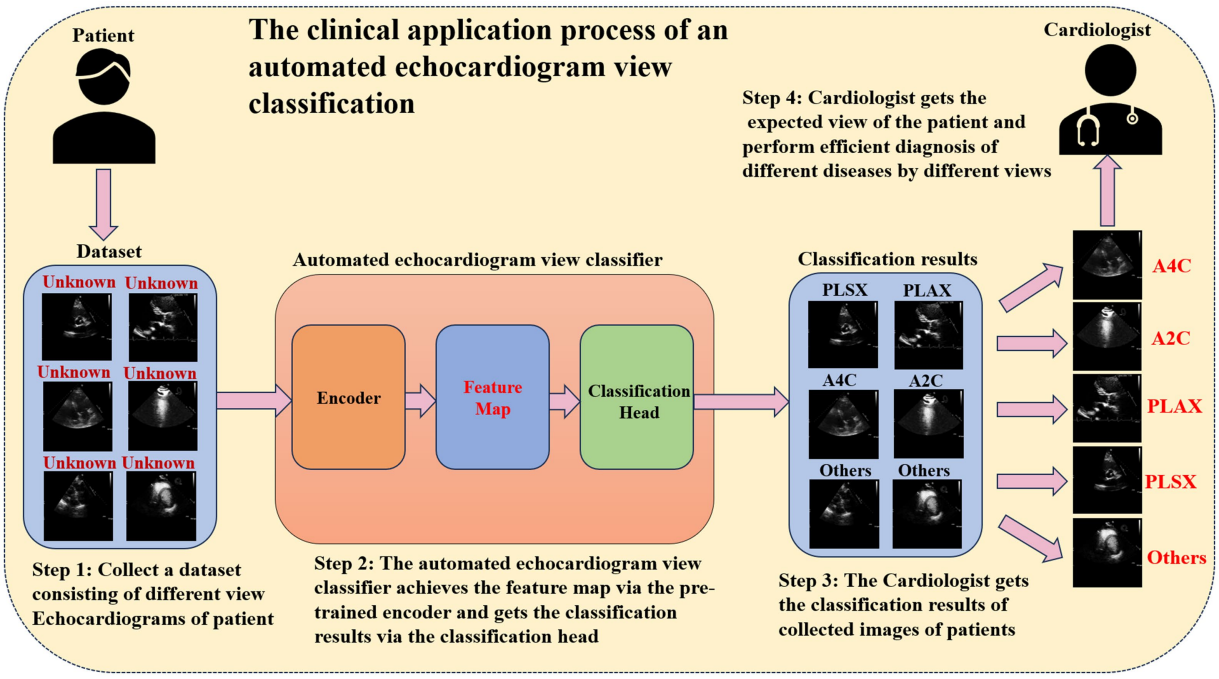

FIGURE 1
The clinical application process of an automated echocardiogram view classification.

classification. Avola et al. (2024) built a multi-scale feature transformer model to capture different scale details to get better classification results in the TMED2 (Huang et al., 2022) dataset. Although this model can provide multi-view and single-view recognition, the transformer model has a more significant number of parameters and lower computational efficiency compared with the lightweight models. To satisfy real medical clinical needs, for the first time, we applied the lightweight model to echocardiogram view classification to improve the classification efficiency. RepViT (Wang et al., 2024) model designed a novel lightweight CNN based on the structures of lightweight VIT and achieved the best performance of the lightweight model. Inspired by the great work, we designed a four-layer lightweight encoder based on the RepViT for view classification. To improve our training efficiency, we decoupled the mixed data generate stage and the pseudo-label generate stage. Specifically, we embedded the DAMix Block into the teacher encoder to provide high-quality mixed data to the classifier. We embedded the SP Block into the student model to leverage the out-of-distribution.

In this paper, we proposed a lightweight open-set semi-supervised method, SPEMix, for echocardiogram view classification. Different from traditional semi-supervised learning classification algorithms, such as FixMatch and OpenMatch, our proposed SPEMix can classify the echocardiogram views accurately and efficiently, providing a new perspective on practical clinical application. FixMatch achieves semi-supervised learning by assigning pseudo-labels to unlabeled data but discards unlabeled data with low thresholds, making it difficult to consider unseen classes. In contrast, our SPEMix assigns open-set pseudo-labels to each unlabeled data point, leveraging the semantic information of each unlabeled instance. Compared to OpenMatch, our open-set semi-supervised method introduces mixup data augmentation, which improves classification accuracy through joint supervision of unseen classes and mixed data. In summary, our contributions are as follows:

1. We innovatively proposed a mixed data generator (DAMix) which introduced the dynamic attention mechanism for the medical echocardiogram. The DAMix Block can provide the mixed mask embedded with the mixing ratio. The mixed mask, which consists of valuable information, can mix up the unlabeled echocardiogram efficiently at the pixel level.
2. We proposed a novel method to leverage the out-of-distribution unlabeled data from superclass probability contribution. We designed an SP Block to model the unlabeled data in terms of superclass distribution to generate the pseudo-labels of unlabeled data. Then, we introduced an open-set classifier to calculate the superclass prediction. We also proposed a novel open-set loss based on consistent regularization to utilize the superclass pseudo-label.
3. Finally, we applied the lightweight model to our semi-supervised method and built a lightweight encoder based on RepViT, which improved the accuracy and efficiency of our proposed method.

2 Methods

In this section, we introduced the detailed information of the proposed novel framework, SPEMix, for echocardiogram view classification. The framework of the proposed SPEMix was shown in Figure 2. The two core components of the SPEMix, DAMix Block and SP Block, can, respectively, generate the mixed mask to mix up data and superclass pseudo-label to leverage the out-of-distribution data. Meanwhile, the SPEMix used the lightweight encoder model to improve efficiency. The lightweight teacher model consisted of the DAMix Block to generate the augmentation data. The lightweight student model consisted of the SP Block to generate the superclass

pseudo-label. The teacher model and student model were trained in an end-to-end manner to enable data enhancement and superclass pseudo-labels generation to work together.

2.1 DAMix block for the generation of high-quality augmented echocardiograms

Mixing augmentation technology (Zhang et al., 2018) is useful for improving the accuracy and generalization of classification tasks. However, the traditional mixing augmentation methods did not work well for echocardiogram view classification because most echocardiograms are grayscale images (Shorten and Khoshgoftaar, 2019). To address these issues and improve the accuracy and generalization of echocardiogram view classification, we proposed DAMix Block to perform pixel-level Efficient Mixup. The DAMix Block generated the mixed mask M that contains crucial information for performing the pixel-level Efficient Mixup process. This DAMix Block was capable of embedding the mixing ratio into the mixed mask. The framework of the DAMix Block was shown in Figure 3.

Inspired by the improvement of the sparse self-attention in the vision transformer (Zhu et al., 2023), the DAMix Block first introduced the dynamic attention mechanism for the echocardiogram. The dynamic attention mechanism divided the echocardiogram into sub-regions and searched for the regions containing critical information in the feature map. This process helped the generated mask to capture the global features of the entire image efficiently. The process of the dynamic attention mechanism was as follows: Initially, the dynamic attention mechanism divided the echocardiogram feature $x \in \mathbb{R}^{\mathrm{H\times W\times C}}$ into S^2 different small square regions, so the size of each region will be $\frac{HW}{S^2}$. We regarded each of the square regions as a token. Each token had $\frac{HW}{S^2}$ features and there were S^2 tokens in total. This process can be achieved by reshaping the feature map into $x' \in \mathbb{R}^{\frac{\mathrm{HW}}{\mathrm{S^2}}\times \mathrm{S^2}\times \mathrm{C}}$. After the divided step, we will get the echocardiogram token features x'. Then, the dynamic attention mechanism calculated the region Query(Q), region Key(K), and region Value(V) of the token features x' by linear projections. The process is shown in Equation 1:

$$\mathrm{Q} = x'\mathrm{W^q}, \mathrm{K} = x'\mathrm{W^k}, \mathrm{V} = x'\mathrm{W^v} \tag{1}$$

Where $\mathrm{W^q}$ was the linear projection weight of the region query, $\mathrm{W^k}$ was the linear projection weight of the region key, and $\mathrm{W^v}$ was the linear projection weight of the region value.

Furthermore, the dynamic attention mechanism calculated the average matrix of the region Query Q, and region Key K. Specifically, this dynamic attention can calculate the average feature of each token. This process can be achieved by performing 2D average pooling with a kernel size of $\frac{HW}{S^2}\times 1$ and reshaping the pooling results into $S^2 \times \mathrm{C}$ as the average matrix $Q^a \in \mathbb{R}^{S^2\times C}$ and $K^a \in R^{S^2\times C}$.

To search for the crucial regions of the echocardiogram efficiently, the similarity between each region was calculated by the result of the dot product of the average matrix of the query and key. The process of calculating the similarity is shown in Equation 2:

$$\mathrm{P} = Q^a\left(K^a\right)^{\mathrm{T}} \tag{2}$$

Where $\mathrm{P} \in \mathbb{R}^{\mathrm{S^2\times S^2}}$ denoted the similarity of the regions; $Q^a \in \mathbb{R}^{S^2\times C}$ denoted the average matrix of the Q, $K^a \in R^{S^2\times C}$ denoted the average matrix of the K. The last step of the dynamic-attention mechanism was to find the index $(id_1, id_2, id_3, \ldots, id_k)$ of the top k most relevant tokens for each token, then we collected all of the relevant tokens of K and V as a new K^c and a new V^c, which can be achieved by searching for the index of P. The dynamic attention can be represented as the Equation 3:

$$output = softmax\left(\frac{Q\left(K^c\right)^T}{\sqrt{c}}\right)V^c \tag{3}$$

DAMix Block was guided by the dynamic attention mechanism to find regions in the echocardiogram that are useful for the classification task. In addition, DAMix Block embedded the mixing rates into the corresponding feature maps and used the idea of cross-attention to generate appropriate mixed masks for the echocardiogram. The process of mixed mask generation can be formulated as follows: getting an unlabeled image pair (u_1, u_2) from the minibatch of the unlabeled dataset. The feature maps from the k-th layer of the unlabeled image pair were inputted into the DAMix Block. The dynamic attention module was subsequently employed to calculate the weighted feature maps z_1', z_2'. Furthermore, to achieve the pixel-level mixup, we designed the Value Block, Query Block, and Key Block to calculate the Value matrix embedded with λ, Query matrix embedded with λ, and Key matrix embedded with λ respectively, where $\lambda \in [0,1]$ is the mixing ratio that satisfies Beta distribution $\lambda - \beta(\alpha,\alpha)$. Specifically, these modules can splice a mixing rate matrix of the same size as the feature over the channel dimension of the feature, where each element of the mixing rate matrix was a randomly generated lambda λ. Then the DAMix Block got the Query matrix Q, Value matrix V, and Key matrix K, respectively, via the query block, value block, and key block. The structure of these blocks were shown in the right part of Figure 3. Notably, the Key block had the same structure as the Query block. To get the mixed mask M, we calculated the similarity matrix P of Q and K by the cross-attention mechanism, and generated the mixed mask M by using the similarity matrix P_2 and Value matrix V. The function of generating M can be formulated as the Equations 4, 5:

$$P_2 = softmax\left(\frac{\mathrm{K}\left(z_1'\right)^T \otimes Q\left(z_2'\right)}{\mathrm{C}}\right) \tag{4}$$

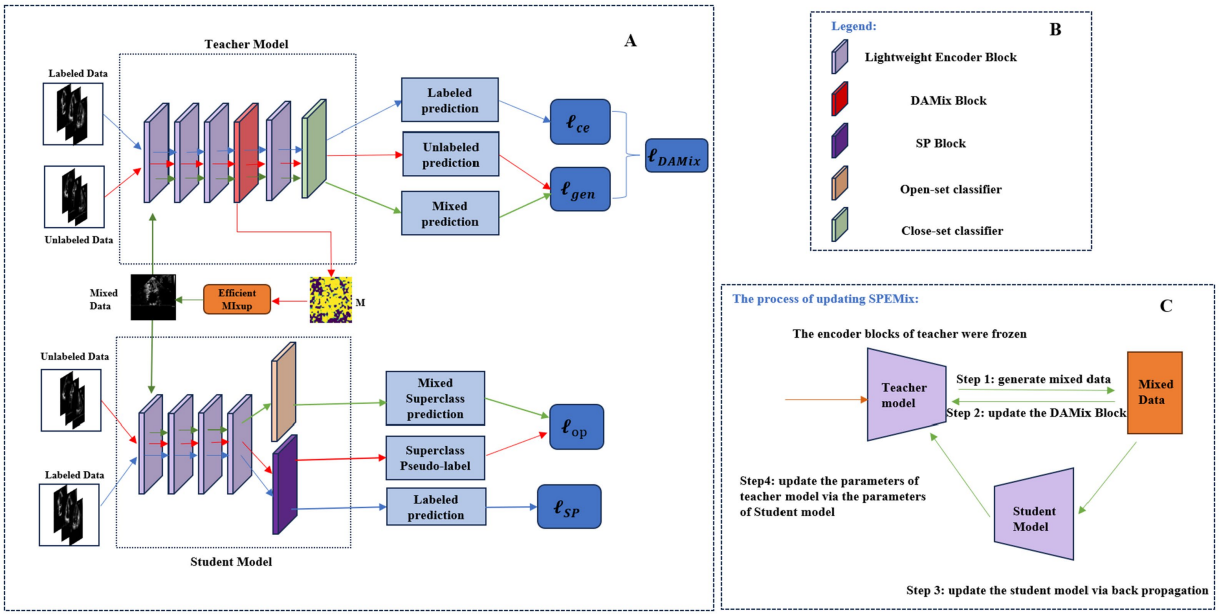

FIGURE 2
The framework of the proposed method **(A)**. Specifically, the DAMix Block was embedded behind the third layer of the teacher model and generated the mixup mask to mix up the input of the unlabeled data. The SP Block was added at the end of the student model to generate the superclass pseudo-label. The open-set classifier can calculate the superclass prediction. The legend of the framework **(A)** is shown in **(B)**. The process of updating SPEMix is shown in **(C)**. The parameters of the student model were updated via backpropagation, and the teacher parameters were updated by the exponential moving average strategy of the student model parameters.

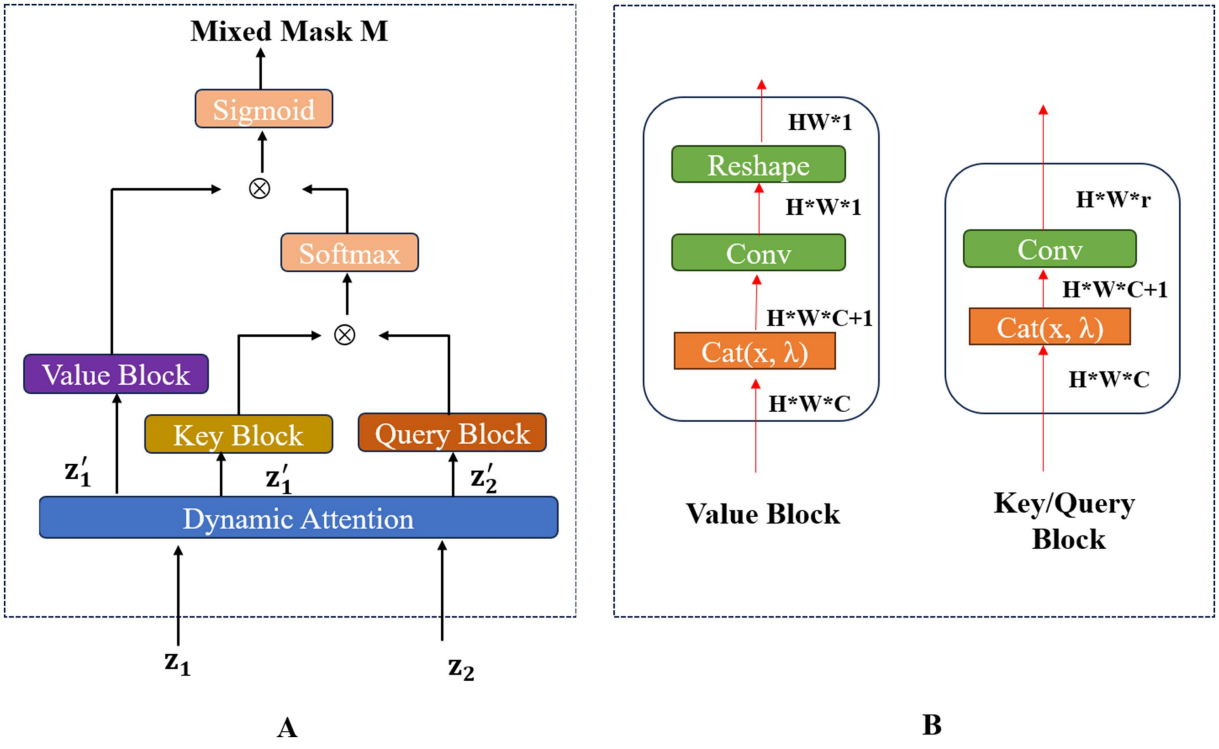

FIGURE 3
The overall framework of the DAMix Block is shown in **(A)**. The structure of the Value Block, Key Block and Query Block is shown in **(B)**.

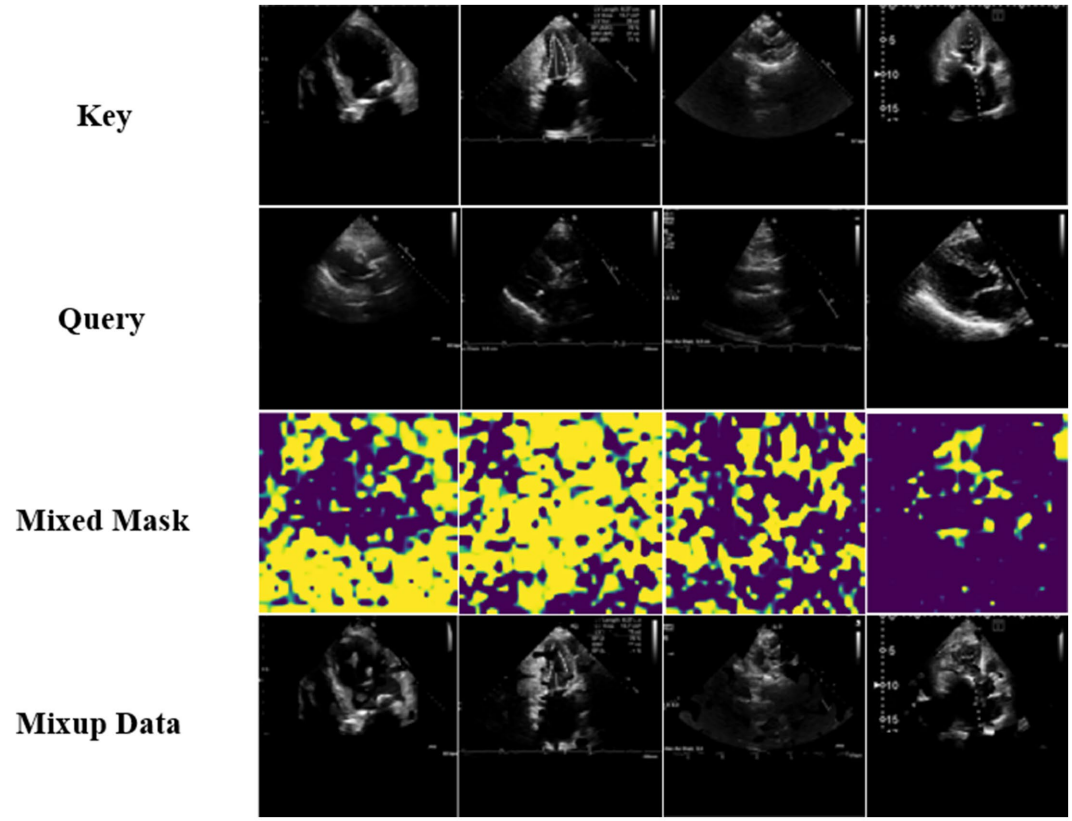

FIGURE 4
Visualization results of the mixed masks generated via DAMix Block and visualization results of mixup data generated via Efficient Mixup.

$$\mathrm{M} = \mathrm{U}\left(\sigma\left(P_2 \otimes \mathrm{V}\left(z_1'\right)\right)\right) \tag{5}$$

Where σ was the Sigmoid activation function, V denoted the value block, K denoted the key block, Q denoted the query block, $\otimes$ denoted the matrix multiplication, and C was a normalization factor, U denoted upsample function.

Based on the mixed masks generated by DAMix Block, we proposed the Efficient Mixup. The Efficient Mixup will generate high-quality mixed images via the mixed mask. Efficient Mixup achieved linear interpolation through pixel-wise multiplication between data and masks, generating high-quality mixed echocardiograms. When performing the efficient mixup between u_1 and u_2, we regard unlabeled image u_1 as the value. The DAMix Block calculated the mixed mask M for the value image. Similarly, the mixed mask of the value image u_2 was 1-M generated via the DAMix Block. The specific formula for efficient mixup was the Equation 6:

$$\text{Efficient Mixup} = M \odot u_1 + (1 - M) \odot u_2 \tag{6}$$

Where $\odot$ denoted element-wise product, u_{mix} was the output result of the DAMix Block. The visualization results of the mixed masks generated via DAMix Block and visualization results of mixup data generated via Efficient Mixup were shown in the Figure 4.

In the training process of our SPEMix, following the previous successful work (Liu et al., 2022), we also embedded our DAMix Block behind the third layer of the encoder. The DAMix loss function can update the parameters of the DAMix Block, a combination of the unlabeled augmentation data generation loss and the labeled data classification loss. To calculate the augmentation generation loss, we proposed the novel generation loss based on the mixed cross-entropy loss function (Liu et al., 2022) for the mixed unlabeled augmentation data. In the previous work, the cross entropy loss between mixed prediction and the mixup of labels was represented by the mixed cross-entropy loss function. Hence, we calculate our mixed generation loss by modifying the mixed cross-entropy. The view of our unlabeled mixed cross-entropy loss function was as follows: we considered the prediction p_1, p_2 of unlabeled data u_1, u_2 before augmentation as the label of the unlabeled data. The generation loss function can be formulated as the Equation 7:

$$\ell_{gen} = \lambda \ell_{ce}\left(p_1, p_{mix}\right) + (1 - \lambda)\ell_{ce}\left(p_2, p_{mix}\right) \tag{7}$$

Where λ meant the mixing ratio, and ℓ_{ce} meant the cross-entropy loss function.

However, the encoder's predictions of the unlabeled data before augmentation were unreliable in the early stages of training, which leaded to the wrong optimization of the generation loss. To address this issue, we also used the supervision of labeled data to assist in the generation of the mixup mask to guide the process of optimizing the

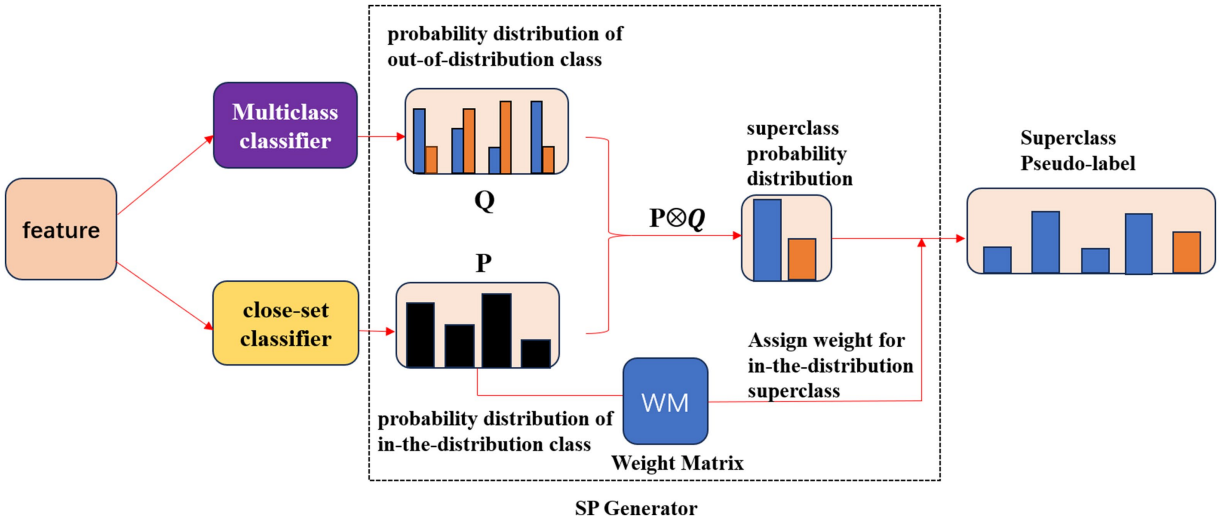

FIGURE 5
The framework of the SP Block. Specifically, the multiclass classifier calculated the probability distribution of out-of-distribution class Q, $Q \in \mathbb{R}^{4\times2}$. The close-set classifier calculated the probability distribution of in-the-distribution class P, $P \in \mathbb{R}^{1\times4}$. Then the SP Generator acquired the superclass probability distribution and generated the Superclass Pseudo-label by assigning the weight matrix for the in-the-distribution superclass.

DAMix Block. The process can be achieved by the cross-entropy of the labeled data, the process can be formulated as the Equation 8:

$$\ell_{labeled} = \ell_{ce}(y, p_x) \tag{8}$$

where y denoted the ground truth of the labeled data, and p_x denoted the prediction of the labeled data. Finally, the loss of the DAMix Block denoted the combination of the generation loss and the cross-entropy loss of the labeled data, be formulated as the Equation 9:

$$\ell_{DAMix} = \ell_{gen} + \ell_{labeled} \tag{9}$$

2.2 SP block for the generation of the echocardiogram superclass pseudo-label

To utilize the unlabeled data of out-of-distribution, we proposed a novel superclass pseudo-label from the perspective of superclass probability. Specifically, our approach first considered all in-the-distribution classes as the in-the-distribution superclass and all out-of-distribution class data as an out-of-distribution superclass. Then, we modeled the superclass from the perspective of the probability distribution to generate the superclass pseudo-label.

We designed the SP Block(short for Superclass Pseudo-label generator) to achieve the above process. The framework of the SP Block is shown in Figure 5. The SP Block included a multiclass classifier to calculate the out-of-distribution class probability distribution, a close-set classifier to calculate the in-the-distribution class probability distribution, and a SP Generator to get the superclass pseudo-label. The black column of P in Figure 5 represented the probability of the sample belonging to each in-the-distribution class only considering the presence of the in-the-distribution classes. The blue columns of Q in Figure 5 represented the probability of the sample belonging to each in-the-distribution class accounting for the unknown classes. Notably, the black column of P was not the same as the blue column of Q. While the orange columns of Q in Figure 5 represented the probability of the sample not belonging to each in-the-distribution class.

Our SP Block generated the superclass pseudo-label for unlabeled samples from the perspective of superclass probability distributions. Our SP Block acquired the in-the-distribution probability distribution P by calculating the probability that the sample belongs to each class in the distribution. Specifically, the close-set classifier, which is a normal fully connected layer, calculated the in-the-distribution probability P, $P \in \mathbb{R}^{1\times4}$. However, we can only get the true class prediction when this unlabeled sample belongs to the in-the-distribution classes. Furthermore, we calculated the out-of-distribution probability distribution Q to model the out-of-distribution classes, which can be calculated by multi-binary classifiers. The multi-binary classifier has been proven to be able to detect whether a sample belongs to each in-the-distribution class,and the technology has been used widely in the previous open semi-supervised learning of filtering the out-of-distribution (Saito et al., 2021). The multiclass classifier consisted of four binary classifiers and each binary classifier predicted the sample whether belonging to the k-th class. Specifically, Ψ_k is a multiclass classifier consisted of k binary classifier $\Psi_k = \{\varphi_1, \varphi_2, \varphi_3, \ldots \varphi_k\}$. φ_k can output $o = (q_k, 1 - q_k)$, where q_k means the probability belongs to the k-th class of the sample. The output of Ψ_k was a matrix of k rows and 2 columns.

Then the SP Generator combined Q and P to generate the superclass pseudo-label, which can be achieved by the matrix multiplication of P and Q. The process of generating the superclass pseudo-label by the SP Generator had three steps: First, the SP Generator calculated the superclass probability distribution $D_p = (D_{in}, D_{out}) \in \mathbb{R}^{1\times2}$, where D_{in} denoted the in-the-distribution superclass probability, D_{out} denoted the out-of-distribution superclass probability. Second, the SP Generator got the in-the-class

probability weight matrix WM of the sample. Finally, the SP Generator assigned weights for the in-the-distribution superclass according to WM and got the superclass pseudo-label. The process of generating a superclass pseudo-label can be formulated as the process of Equations 10–12:

$$D_p = P \otimes Q \tag{10}$$

$$\mathrm{WM} = (\alpha_1, \alpha_2, \ldots \alpha_n) = \left(\frac{p_1}{\sum_{i=1}^{n} p_i}, \frac{p_2}{\sum_{i=1}^{n} p_i}, \ldots, \frac{p_i}{\sum_{i=1}^{n} p_i} \right) \tag{11}$$

$$\mathrm{SP} = \left(\mathrm{WM} D_{in}, D_{out} \right) \tag{12}$$

Where $\mathrm{WM} \in R^{l \times n}$, where n denoted the number of the in-the-distribution class. SP denoted the superclass pseudo-label. p_i denoted the i-th element of the in-the-distribution probability P.

To optimize the SP Block, we, respectively, optimized the multiclass classifier and the close-set classifier with labeled data. For the loss of the multiclass classifier, we used the hard-negative sampling strategy (Saito and Saenko, 2021), following the previous work. The loss function of the multiclass classifier can be formulated as the Equation 13:

$$\ell_{mul} = \frac{1}{B} \sum_{i=1}^{B} \left(-\log\left(p_{i,y_i} \right) - \min_{k \neq y_i} \log\left(\overline{p_k} \right) \right) \tag{13}$$

Where B represented the batch size of the labeled data, $p_{i,y}$ represented the first element in the y-th row of the out-of-distribution probability Q. $\overline{p_k}$ represented the second element in the k-th row of the out-of-distribution probability Q.

For the loss of the close-set classifier, we used the cross-entropy loss function of the labeled data to optimize, was shown in the Equation 14:

$$\ell_{close} = \ell_{ce}\left(y, p_x \right) \tag{14}$$

$$\ell_{SP} = \ell_{close} + \ell_{mul} \tag{15}$$

The total loss of the SP Block was shown in Equation 15. In our SPEMix, we hoped the proposed DAMix Block and SP Block could work together to utilize the unlabeled datasets efficiently. Specifically, the SP Block can assign the superclass pseudo-label for augmentation data generated by the DAMix Block. To achieve the process, we built an open-set classifier to predict the superclass pseudo-label. We proposed the open-set loss function of SP Block based on the mixed process of the superclass pseudo-label to utilize the augmentation of unlabeled images. Specifically, the processes of calculating the open-set loss function are shown in the Equations 16–18:

$$\ell_{op} = \lambda \ell_{ce}\left(\mathrm{y}_{spu1}, o_{mix} \right) + (1 - \lambda) \ell_{ce}\left(\mathrm{y}_{spu2}, o_{mix} \right) \tag{16}$$

$$\mathrm{y}_{spu1} = \max\left(SP_{u1} \right) \tag{17}$$

$$\mathrm{y}_{spu2} = \max\left(SP_{u2} \right) \tag{18}$$

Where SP_{u1} meant the superclass pseudo-label of u_1, SP_{u2} meant the superclass pseudo-label of u_2. y_{spu2} meant the corresponding class of the pseudo-label SP_{u2}. y_{spu1} meant the corresponding class of the pseudo-label SP_{u1}. max was the function to caculate the index of the maximum probability. o_{mix} denoted the superclass pseudo-label prediction of augmentation image from u_1 and u_2 via the open-set classifier.

2.3 Lightweight encoder and end-to-end efficient learning paradigm

For the first time, we applied the lightweight model to the echocardiogram view classification. Specifically, we designed a lightweight encoder based on the RepViT (Wang et al., 2024) network for our SPEMix. RepViT implemented the design of lightweight networks by using the structural re-parameterization (Ding et al., 2019) principle. Hence, we built our lightweight encoder following the structure of RepViT to improve the efficiency of echocardiogram view classification. We built the lightweight student encoder and lightweight teacher encoder for our SPEMix. To satisfy our view classification tasks, our RepViT lightweight encoder only had four blocks.

Inspired by the success of AutoMix (Liu et al., 2022), we also adopted the momentum update pipeline to decouple the process of augmentation of unlabeled images and the process of assigning the superclass pseudo-label. The architecture was shown in Figure 2. We constructed two lightweight encoders with identical initialized parameters, which can help SPEMix synchronize the two processes by employing end-to-end training and achieve better accuracy and generalization. Specifically, the DAMix Block in the teacher model mixed up the unlabeled data, and the SP Block in the student model generated superclass pseudo labels for unlabeled data. The parameters of the student model can be updated by back propagation, while the teacher parameters were updated by the exponential moving average strategy (Polyak and Juditsky, 1992) from the parameters of the student model. The total loss function of our SPEMix was the Equation 19:

$$\ell_{total} = \ell_{DAMix} + \ell_{SP} + \ell_{op} \tag{19}$$

The process of updating the teacher model can be reformulated as the Equation 20:

$$\theta_{teacher} = m\theta_{teacher} + (1 - m)\theta_{student} \tag{20}$$

where m was the momentum coefficient and $m \epsilon (0,1], \theta_{student}$ denoted the parameters of the student model, $\theta_{teacher}$ denoted the parameters of teacher model.

3 Results and discussion

3.1 Implement details

3.1.1 Dataset

We used the *Tufts Medical Echocardiogram Dataset 2* (TMED2) (Huang et al., 2022) to train the proposed SPEMix and the CAMUS (Leclerc et al., 2019) dataset and Unity (Howard et al., 2021) dataset

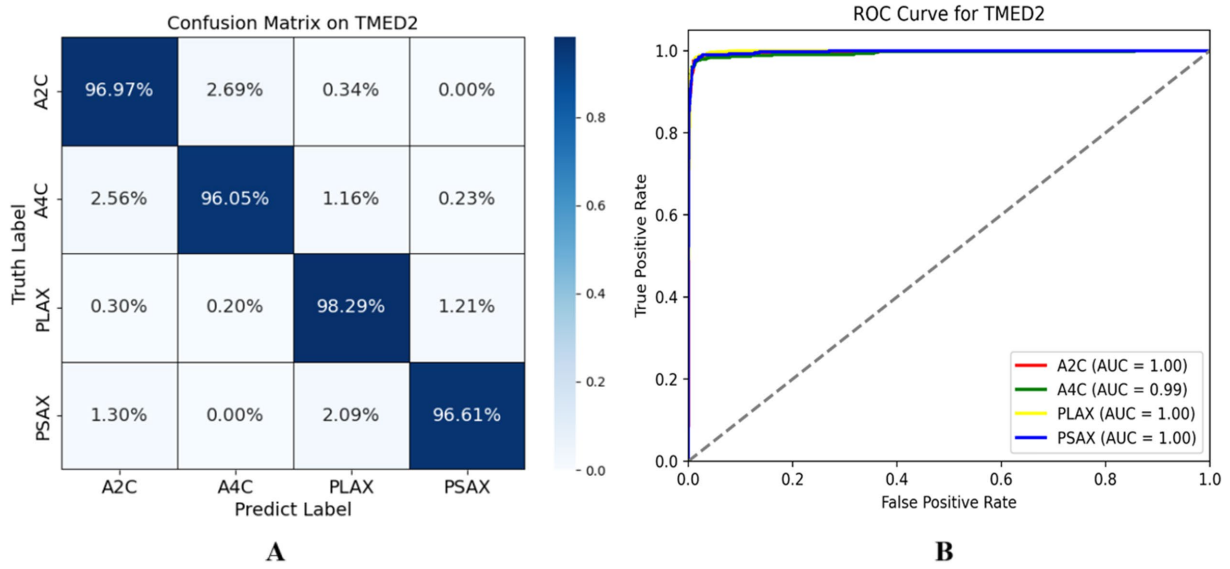

FIGURE 6
The evaluation result of the SPEMix. The left diagram has some problems due to decimal retention, so we provide a new confusion matrix on TMED2 **(A)** represents the confusion matrix of SPEMix on TMED2 and the right diagram **(B)** represents the ROC curve of TMED2.

to evaluate the generalization. Specifically, the TMED2 contains four types of echocardiogram views, including PLAX, PLSX, A2C, and A4C. This dataset provides 353,500 unlabeled images and 24,964 labeled data collected from different patients. For TMED2, we used the officially released training sets, test sets, and validation sets. All of the resolution of this dataset is 112×112 pixels. We used the RandomCrop and RamdomHorizontalFlip as basic augments for the training dataset during training. The CAMUS (Leclerc et al., 2019) dataset contains two view types including A2C, and A4C. We resized the resolution to 112×112 pixels to evaluate the generalization of the classification. The Unity (Howard et al., 2021) dataset contains three view types, including PLAX, A2C, and A4C. We resized the resolution to 112×112 pixels to evaluate the generalization of the classification.

3.1.2 Training setting

For training our SPEMix, we used our designed lightweight encoder with 112×112 size inputs. For the hyper-parameter of the SPEMix, the momentum coefficient was set to 0.999. For labeled data batch size and unlabeled data batch size, we set them to 64. We used the Adam optimizer to update the model parameters. For learning rate, we chose from the set of {0.1,0.01,0.001,0.0001}, and we reported the different best learning rate in different experiments. The learning rate of the SPEMix was set to 0.0001, we trained the SPEMix 500 epochs and adapted the learning rate by the Cosine Schedule (Loshchilov and Hutter, 2022). We did not use the warm-up strategy. During the process of training, the parameters of the student model were updated via back propagation. And the parameters of the DAMix Block can update via the loss function of DAMix. The parameters of the DAMix Block will be frozen when the parameters of the teacher model update based on the parameters of student model. For comparison experiments, we used the Adam optimizers to train the other methods and we chose the best learning rate for each method. All of the comparison methods were trained on the TMED2. To make a fair comparison, all of our experiments were implemented on the GPU of the model NVIDIA A40.

3.2 Results of SPEMix

We reported the performance of our lightweight encoder via SPEMix on the TMED2 test dataset, the confusion matrix of the test dataset was shown in the left diagram of Figure 6. Each element of the matrix represented the probability of being predicted as the corresponding class, and the diagonal value represented the prediction accuracy of each view. The classification accuracy of the A2C view reached 96.97%, the classification accuracy of A4C reached 96.05%, the PLAX classification accuracy reached 98.29%, and the PSAX classification reached 96.61%. These results demonstrated that our SPEMix predicted every view very well. At the same time, we also provided the ROC curve to evaluate our classifier. We gave the ROC curve of SPEMix and calculated the AUC for each class. The larger values of the AUC mean the better the performance of our classifier.

3.3 Comparison results with mixing methods

In this section, we compared our proposed SPEMix with the previous SOTA mixup methods to prove the advanced performance of our SPEMix, including CutMix (Yun et al., 2019), SaliencyMix (Uddin et al., 2006), and AutoMix (Liu et al., 2022). CutMix represents the typical mixup method in natural images. SaliencyMix is the SOTA mixup that can generate mixed data by utilizing the saliency information of the natural images. AutoMix is the SOTA of the pixel-level mixup method which gets the most advanced performance in

TABLE 1 Comparison results between SPEMix and the previous Sota mixing methods. All results are expressed as percentages (%).

Method	Encoder	Test accuracy
CutMix (Yun et al., 2019)	Wideresnet (Zagoruyko and Komodakis, 2016)	94.30 ± 0.2
CutMix (Yun et al., 2019)	Lightweight encoder(Ours)	95.97 ± 0.67
SaliencyMix (Uddin et al., 2006)	Wideresnet (Zagoruyko and Komodakis, 2016)	96.31 ± 0.43
SaliencyMix (Uddin et al., 2006)	Lightweight encoder(Ours)	96.87 ± 0.2
AutoMix (Liu et al., 2022)	Wideresnet (Zagoruyko and Komodakis, 2016)	96.34 ± 0.19
AutoMix (Liu et al., 2022)	Lightweight encoder(Ours)	96.91 ± 0.27
SPEMix(Ours)	Wideresnet (Zagoruyko and Komodakis, 2016)	**97.21 ± 0.10**
SPEMix(Ours)	Lightweight encoder(Ours)	**97.28 ± 0.11**

To make a fair comparison, we ran each method with three trials and reported the mean and standard deviation of test accuracy on the TMED2 dataset. The best performance of each encoder was highlighted in bold.

natural images. To compare these methods fairly, we, respectively, trained a common WideResNet encoder classifier and our lightweight encoder classifier by using each mixup method. We chose the best parameters for all of the methods in this comparison experiments. The best learning rate was set to 0.01. The batch size was set to 64, we all used the Cosine Schedule to adapt the learning rate and the training epoch was set to 500. All of the experiments did not use the warm-up strategy. The comparison result was shown in Table 1. We reported the mean accuracy and standard deviation on the TMED2 from three different trails.

To better investigate the classification process of each method, we visualized the performance of each method via the technique of CAM (Selvaraju et al., 2017). The visual results of the comparison experiment are shown in Figure 7. From the visual result, we can intuitively observe that the proposed SPEMix is more adept at focusing on crucial information compared to other advanced data augmentation methods. Based on the comparison experiment results in Table 1 and Figure 7, we got the following conclusions:

1. Our proposed SPEMix is less affected by the echocardiogram's background information than CutMix. Our SPEMix can improve by 2.91% with the encoder of Wideresnet and improve by 1.31% with the lightweight encoder compared to CutMix. To better understand the experimental results, we compared the visualization results of CutMix and SPEMix in Figure 7. These results illustrated that CutMix focuses more on the background information of the echocardiogram. This is mainly because cutmix performs mixup augmentation by randomly selecting regions of the image. This approach results in the appearance of augmentation data containing only background information, leading to the trained model not differentiating well between the foreground and background regions of a medical image. Hence, the classification performances of Cutmix are more affected by irrelevant information.
2. Our proposed SPEMix focused on more regions containing salient echocardiogram information than SaliencyMix. Our SPEMix improved by about 0.9% with the encoder of Wideresnet and about 0.41% with the encoder of RepViT compared to SaliencyMix. The saliency information helped our SPEMix and SaliencyMix focus on the vital information from the visual results in Figure 7. However, the SPEMix focused more regions on the salient information of the echocardiographic foreground region when performing view classification. The main reason is that our SPEMix can include more salient details in the foreground by seeking detailed salient information at the pixel level. However, SaliencyMix seeks salient areas at the image level, and this method cannot consider the detailed features.
3. Our proposed SPEMix sought vital information more easily from an echocardiogram than Automix. Our SPEMix can improve by 0.87% with the encoder of Wideresnet and by 0.37% with the encoder RepViT compared to AutoMix. These visualization results illustrated that SPEMix more easily focused on vital information. These results were mainly because our SPEMix generated the mixed mask with the assistance of dynamic attention. The mixed masks containing vital information helped the model easily seek important details. However, Automix generated the mixed mask only considering the similarity of the two images. The background information of the medical images was also similar between different views. This background information prevented the model from seeking vital information.
4. Furthermore, the experimental results also revealed the potential of the designed lightweight network for application in different methods. The designed lightweight network improved the performance of TMED2 compared to the traditional WideResnet encoder using various techniques for training. In CutMix, SaliencyMix, and AutoMix, the lightweight network enhanced by 1.67%, 0.56%, and 0.57%, respectively. Using our SPEMix method for training, the lightweight network still maintains better performance. The experimental results demonstrate the greater feasibility of lightweight networks in view classification.

In summary, our SPEMix can achieve the best test classification accuracy on the WideResNet encoder and our lightweight encoder compared to other mixup methods. These experiment results demonstrate our SPEMix has advanced performance in view classification tasks.

3.4 Comparison results with semi-supervised learning

In this section, we compared the performance of SPEMix with other competitive semi-supervised learning methods, which

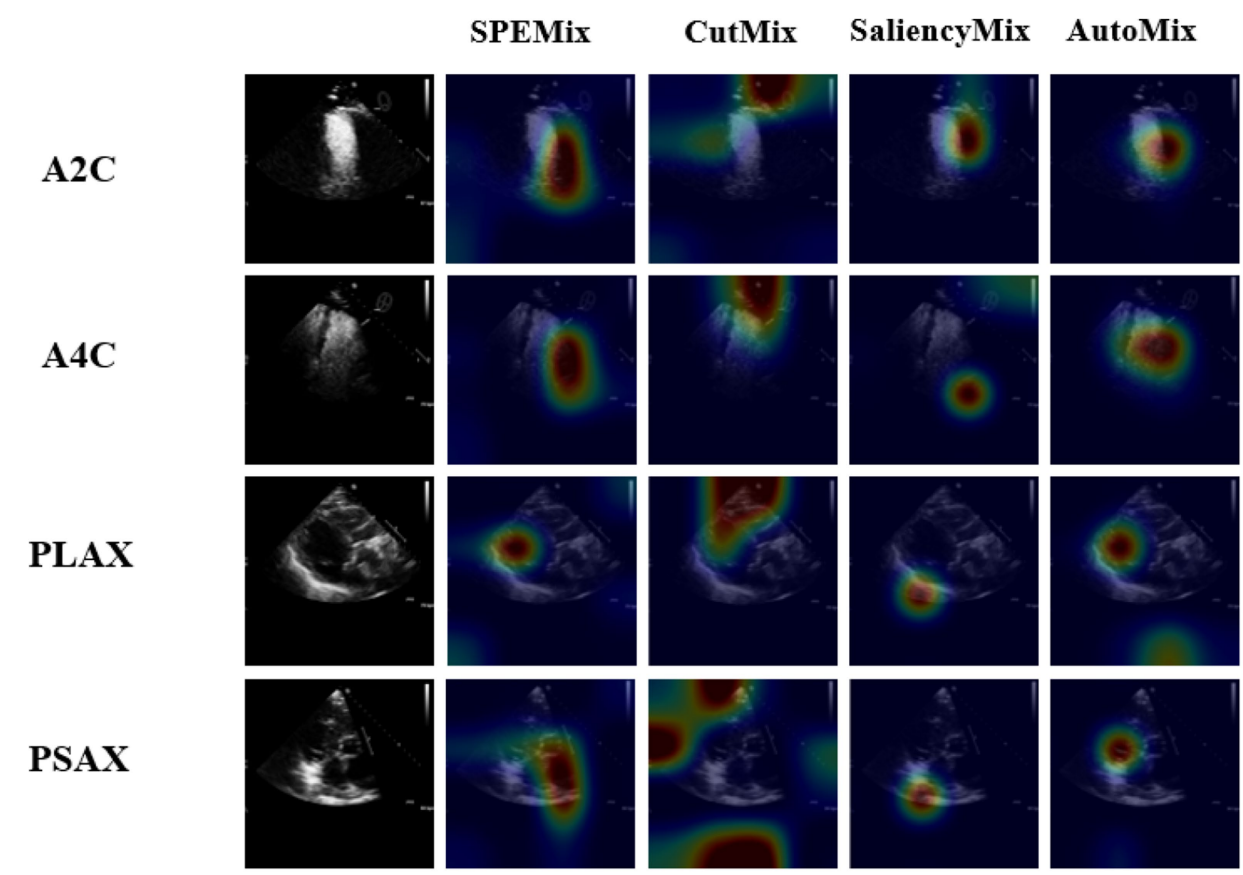

FIGURE 7
The class activation mapping (CAM) for classifiers that are trained based on different mixing methods. The data of each view is chosen from the validation datasets randomly. The visualization results of regions focused on lightweight models trained by different methods. Our proposed SPEMix can better focus on the vital regions than other SOTA methods.

TABLE 2 The performance of different semi-supervised methods on three datasets. All results are expressed as percentages (%).

Methods	TMED2 (Huang et al., 2022)	CAMUS (Leclerc et al., 2019)	Unity (Howard et al., 2021)
FixMatch (Sohn et al., 2020)	95.49 ± 0.07	81.39 ± 0.72	91.41 ± 1.28
$Fix-a-step^{\dagger}$ (Huang et al., 2023)	95.64 ± 0.16	83.20 ± 0.74	93.72 ± 0.65
OpenMatch (Saito et al., 2021)	96.31 ± 0.06	83.37 ± 2.42	92.11 ± 1.08
IntereLUDE* (Huang et al., 2024)	96.55 ± 0.39	86.25 ± 6.22	**96.14 ± 0.48**
IntereLUDE+* (Huang et al., 2024)	96.75 ± 0.17	81.88 ± 8.37	94.47 ± 0.85
SPEMix(Ours)	**97.28 ± 0.11**	**87.64 ± 0.67**	94.71 ± 0.24

[†]Means the method follows the parameters of the cited work with the encoder of Wideresnet and uses SGD optimizer. *Means the results cited from the SOTA work. Other methods used our lightweight encoder and the Adam optimizer. The best performances of different datasets in bold.

include FixMatch (Sohn et al., 2020), Fix-a-step (Huang et al., 2023), OpenMatch (Saito et al., 2021), and InterLUDE (Huang et al., 2024). We test different methods on TMED2, CAMUS, and Unity datasets to evaluate the accuracy and generalization of these methods. We trained the FixMatch and OpenMatch with the optimal parameters. Specifically, each method was trained with our lightweight encoder, the Adam optimizer, and 500 epochs. The learning rate was set to 0.001. We also trained Fix-a-step with the optimal parameters(Wideresnet as encoder and SGD optimizer) given in their paper to ensure that their method can achieve the most accurate classification results. For the InterLUDE and IntereLUDE+(IntereLUDE with Self-Adaptive Threshold and Self-Adaptive Fairness), we directly quote the results given in the paper(the code is not yet open source). The comparison results for accuracy and training time are presented in Tables 2, 3, respectively.

From the comparison results, we found that our SPEMix can achieve better accuracy on the three datasets. Compared to the FixMatch, the accuracy, and generalization of our SPEMix are improved by 1.79%, 6.25%, and 3.3% on TMED2, CAMUS, and Unity, respectively. The reason is that the proposed SP Block can leverage the out-of-distribution data to improve the accuracy. FixMatch focuses on the close-set problem and does not pay attention to the

TABLE 3 The comparison results of training time between the traditional semi-supervised method and SPEMix.

Method	Total time	Average time
OpenMatch	4595.57 s	15.32 s
FixMatch	2746.83 s	9.16 s
SPEMix(Ours)	2714.18 s	9.05 s

TABLE 4 The comparison results of different encoders in our comparison experiments. All accuracies are expressed as percentages (%).

Encoder model	Parameters	Accuracy
WideResNet-28-2 (Zagoruyko and Komodakis, 2016)	5.93 M	97.24
Lightweight encoder(Ours)	0.70 M	97.34

out-of-distribution echocardiograms so FixMatch has worse results and is not suitable for open-set tasks (Saito et al., 2021). Meanwhile compared to OpenMatch, the accuracy of our SPEMix is improved by 0.97%, 4.27%, and 2.6% on TMED2, CAMUS, and Unity, respectively. The reason is that our SPEMix can also gain more improvement from the unlabeled datasets via the proposed mixed unlabeled consistency regularization. However, OpenMatch filtered out out-of-distribution data during the training process via its consistency regularization to process the open set data. In this way, OpenMatch neglected some vital information of out-of-distribution so that it has worse results. Compared to the advanced method for similar medical tasks, Fix-a-step, the SPEMix also improves the classification accuracy on different three datasets. Specifically, the accuracy and generalization of SPEMix compared to Fix-a-step improved by 1.64%, 4.44%, and 0.99% on TMED2, CAMUS and Unity, respectively. The reason for these results is that our SPEMix modeled the out-of-distribution data using superclass distribution to utilize all of the unlabeled data efficiently. However, Fix-a-step only used limited unlabeled data, which can improve the classification accuracy. In this way, Fix-a-step only got finite information from the limited out-of-distribution data. Simultaneously, our SPEMix also compares recent SOTA semi-supervised methods in view classification tasks, IntereLUDE and IntereLUDE+(IntereLUDE with Self-Adaptive Threshold and Self-Adaptive Fairness). Compared to the IntereLUDE, our SPEMix can improve by 0.73% and 1.39% on TMED2 and CAMUS. Compared to the IntereLUDE+, our SPEMix can improve by 0.53%, 5.76%, and 0.24% on TMED2, CAMUS, and Unity, respectively. SPEMix has less accuracy in the Unity dataset than IntereLUDE. However, IntereLUDE evaluated the accuracy on the widereset and did not explore a lightweight model. Therefore, our SPEMix still outperforms IntereLUDE overall.

These demonstrate that the accuracy and generalization of our proposed SPEMix outperform the SOTA methods in the view classification task. In summary, all of these results indicate the superior performance and generalization ability of the proposed SPEMix.

3.5 Comparison results between different encoders

In this section, we explored the performance of our proposed lightweight encoder. We reported the number of parameters of the different encoders used in our comparison experiments of Section 3.3, i.e., the number of the parameters of the WideResNet-28-2 and our proposed lightweight encoder. We also reported the accuracy of the two encoders on the TMED2 test dataset after training via SPEMix. The comparison results were shown in Table 4.

The results in Table 4 demonstrated that the parameter number of our proposed encoder is reduced by 88.2% compared with the parameter number of the WideResNet-28-2. Additionally, the results also demonstrated that our proposed lightweight encoder can also maintain the classification performance compared with the WideResNet-28-2 when training via the proposed SPEMix. Otherwise, these results of Table 1 also denoted that, with different mixing methods, our proposed lightweight encoder was better than the WideResNet-28-2. These results demonstrated that our proposed SPEMix got better performance with fewer parameters. Hence, our proposed method had the potential for clinical application.

3.6 Results of ablation experiments

The proposed SPEMix included two core components, DAMix and SP Block, to perform the mixed data augmentation and superclass pseudo-label generation, respectively. In order to explore the effect of each component in SPEMix, the lightweight network was regarded as the baseline in the ablation experiment, and we added DAMix Block and SP Block to the baseline step by step. The final results of the ablation experiments were in Table 5.

Through the results of the ablation experiment, we can observe that the DAMix Block improved the baseline by 1.82%, 6.63%, and 6.04% on TMED2, CAMUS, and Unity. These results demonstrated the Efficient Mixup based on our proposed DAMix Block can improve accuracy. Furthermore, the SP Block improved the baseline by 2.01%, 0.73%, and 8.29% on TMED2, CAMUS, and Unity. The results denoted the advancement of utilizing the out-of-distribution data from the superclass probability perspective. The SPEMix can increase the baseline by 3.14%, 12.18%, and 9.35% on TMED2, CAMUS, and Unity, which illustrates the SPEMix can fuse the advantages of the DAMix Block and SP Block. These results demonstrated the effectiveness of the proposed SPEMix. The DAMix Block can generate corresponding mixed masks embedded with a mixing ratio for unlabelled data. The high-quality mixed data generated through the mixed masks can improve classification accuracy and generalization. SP Block can assign superclass pseudo-labels to unlabelled data through the perspective of superclass distribution to make use of the out-of-distribution data. This out-of-distribution information improved the classification accuracy and generilization. SP Block used an end-to-end approach to fuse two phases. Firstly, the unlabelled mixed data was generated using DAMix, while the information of generated mixed data was leveraged by assigning superclass pseudo-labels through SP Block. The high-quality mixed data generated by DAMix can also enrich the information of out-of-distribution that be leveraged by the SP Block.

4 Conclusion

In this work, we proposed a novel lightweight open-set semi-supervised learning method, SPEMix, to improve the accuracy and generalization of the echocardiogram view classification. The proposed

TABLE 5 The results of the ablation experiment. All results are expressed as percentages (%).

	TMED2 (Huang et al., 2022)	CAMUS (Leclerc et al., 2019)	Unity (Howard et al., 2021)
Baseline	94.2	74.93	85.07
Baseline+DAMix	96.02	81.56	91.11
Baseline+SP Block	96.21	82.56	93.36
SPEMix	97.34	87.11	94.42

The baseline was the proposed lightweight encoder. We trained the baseline by using the TMED2 and reported the classification accuracy on different test datasets.

DAMix Block generated the masks embedded with the mixing ratio containing vital information. These masks efficiently generated high-quality mixed data at the pixel level. Then, the proposed SP Block can generate the superclass pseudo-label from the superclass probability perspective to utilize the vital information from the unlabeled medical dataset. In this way, the proposed SP Block used the out-of-distribution data more effectively. Meanwhile, a novel loss function based on unlabeled consistent regularization was proposed to make the classification model better optimized from the supervision of the mixed unlabeled data and the super pseudo-label. Otherwise, we built a lightweight encoder based on RepViT to decrease the model parameters and improve the classification efficiency. Experiment results indicated that our proposed SPEMix achieved better performance and generalization than other semi-supervised learning methods. Our SPEMix had the potential for clinical application. Although the proposed SPEMix method shows encouraging results for echocardiogram view classification, there are still several areas that require further investigation. Future research could aim to adapt the method for more complex or multi-modal medical datasets, incorporating additional imaging modalities (such as CT or MRI) and patient metadata into the semi-supervised framework.

Data availability statement

The Tufts Medical Echocardiogram Dataset 2 (TMED2) (Huang et al., 2022) was used to train the proposed SPEMix. The CAMUS (Leclerc et al., 2019) dataset and Unity (Howard et al., 2021) dataset were used to evaluate the generalization.

Author contributions

SM: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. YZ: Data curation, Methodology, Software, Writing – review & editing. DL: Methodology, Software, Writing – review & editing. YS: Formal analysis, Writing – review & editing. ZQ: Validation, Funding acquisition, Writing – review & editing. LW: Validation, Funding acquisition, Writing – review & editing. SD: Conceptualization, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Natural Science Foundation of China under Grant (no. 62202092), the Fundamental Research Funds for the Central Universities (no. 2572021BH03), Heilongjiang Provincial Key Research and Development Plan 2023ZX02C10, 2022ZX01A30, and GA23C007, Hunan Provincial Key Research and Development Plan 2023SK2060, Jiangsu Provincial Key Research and Development Plan BE2023081. All the above funders provide the data access support for this research.

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.

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

Avola, D., Cannistraci, I., Cascio, M., Cinque, L., Fagioli, A., Foresti, G. L., et al. (2024). MV-MS-FETE: multi-view multi-scale feature extractor and transformer encoder for stenosis recognition in echocardiograms. *Comput. Methods Prog. Biomed.* 245:108037. doi: 10.1016/j.cmpb.2024.108037

Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., et al. (2017). *Semi-supervised learning for network-based cardiac MR image segmentation*. Medical Image Computing and Computer-Assisted Intervention– MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part II 20; Springer.

Calderon-Ramirez, S., Yang, S., and Elizondo, D. (2022). Semisupervised deep learning for image classification with distribution mismatch: a survey. *IEEE Trans. Artif. Intell.* 3, 1015–1029. doi: 10.1109/TAI.2022.3196326

Chebli, A., Djebbar, A., and Marouani, H. F. (2018). *Semi-supervised learning for medical application: A survey*. 2018 international conference on applied smart systems (ICASS), IEEE.

Degerli, A., Kiranyaz, S., Hamid, T., Mazhar, R., and Gabbouj, M. (2024). Early myocardial infarction detection over multi-view echocardiography. *Biomed. Signal Proc. Control* 87:105448. doi: 10.1016/j.bspc.2023.105448

Ding, X., Guo, Y., Ding, G., and Han, J. (2019). *Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks*. Proceedings of the IEEE/CVF international conference on computer vision.

Gao, X., Li, W., Loomes, M., and Wang, L. (2017). A fused deep learning architecture for viewpoint classification of echocardiography. *Inf. Fusion* 36, 103–113. doi: 10.1016/j.inffus.2016.11.007

Hagberg, E., Hagerman, D., Johansson, R., Hosseini, N., Liu, J., Björnsson, E., et al. (2022). Semi-supervised learning with natural language processing for right ventricle classification in echocardiography—a scalable approach. *Comput. Biol. Med.* 143:105282. doi: 10.1016/j.compbiomed.2022.105282

Han, Y., Duan, B., Guan, R., Yang, G., and Zhen, Z. (2024). LUFFD-YOLO: a lightweight model for UAV remote sensing Forest fire detection based on attention mechanism and multi-level feature fusion. *Remote Sens.* 16:2177. doi: 10.3390/rs16122177

Han, Y., Guo, J., Yang, H., Guan, R., and Zhang, T. (2024). SSMA-YOLO: a lightweight YOLO model with enhanced feature extraction and fusion capabilities for drone-aerial ship image detection. *Drones* 8:145. doi: 10.3390/drones8040145

Howard, J. P., Stowell, C. C., Cole, G. D., Ananthan, K., Demetrescu, C. D., Pearce, K., et al. (2021). Automated left ventricular dimension assessment using artificial intelligence developed and validated by a UK-wide collaborative. *Circ. Cardiovasc. Imaging* 14:e011951. doi: 10.1161/CIRCIMAGING.120.011951

Huang, Z., Long, G., Wessler, B., and Hughes, M. C. *A new semi-supervised learning benchmark for classifying view and diagnosing aortic stenosis from echocardiograms*. Machine Learning for Healthcare Conference, PMLR. (2021).

Huang, Z., Long, G., Wessler, B., and Hughes, M. C. (2022). *TMED 2: A dataset for semi-supervised classification of echocardiograms*. DataPerf: Benchmarking Data for Data-Centric AI Workshop.

Huang, Z., Sidhom, M. J., Wessler, B., and Hughes, M. C. (2023). *Fix-A-step: semi-supervised learning from Uncurated unlabeled data*. International Conference on Artificial Intelligence and Statistics, PMLR.

Huang, Z., Yu, X., Zhu, D., and Hughes, M. C. (2024). InterLUDE: interactions between labeled and unlabeled data to enhance semi-supervised learning. *arXiv* 2024:240310658. doi: 10.48550/arXiv.2403.10658

Kusunose, K., Haga, A., Inoue, M., Fukuda, D., Yamada, H., and Sata, M. (2020). Clinically feasible and accurate view classification of echocardiographic images using deep learning. *Biomol. Ther.* 10:665. doi: 10.3390/biom10050665

Kwon, H. (2023). Adversarial image perturbations with distortions weighted by color on deep neural networks. *Multimed. Tools Appl.* 82, 13779–13795. doi: 10.1007/s11042-022-12941-w

Kwon, H., and Lee, S. (2023). Detecting textual adversarial examples through text modification on text classification systems. *Appl. Intell.* 53, 19161–19185. doi: 10.1007/s10489-022-03313-w

Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., et al. (2019). Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans. Med. Imaging* 38, 2198–2210. doi: 10.1109/TMI.2019.2900516

Lee, J., Kim, T., Bang, S., Oh, S., and Kwon, H. (2024). Evasion attacks on deep learning-based helicopter recognition systems. *J Sens* 2024, 1–9. doi: 10.1155/2024/1124598

Li, Z., Qi, L., Shi, Y., and Gao, Y. (2023). *IOMatch: Simplifying open-set semi-supervised learning with joint inliers and outliers utilization*. Proceedings of the IEEE/CVF International Conference on Computer Vision.

Liu, Z., Li, S., Wu, D., Liu, Z., Chen, Z., Wu, L., et al. (2022). *Automix: Unveiling the power of mixup for stronger classifiers*. European Conference on Computer Vision, Springer.

Loshchilov, I., and Hutter, F. (2022). *SGDR: Stochastic gradient descent with warm restarts*. International conference on learning representations.

Madani, A., Arnaout, R., Mofrad, M., and Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit. Med.* 1:6. doi: 10.1038/s41746-017-0013-1

Madani, A., Ong, J. R., Tibrewal, A., and Mofrad, M. R. (2018). Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit. Med.* 1, 1–11. doi: 10.1038/s41746-018-0065-x

Oliver, A., Odena, A., Raffel, C. A., Cubuk, E. D., and Goodfellow, I. (2018). Realistic evaluation of deep semi-supervised learning algorithms. *Adv. Neural Inf. Proces. Syst.* 31:11.

Polyak, B. T., and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control. Optim.* 30, 838–855. doi: 10.1137/0330046

Saito, K., Kim, D., and Saenko, K. (2021). Openmatch: open-set semi-supervised learning with open-set consistency regularization. *Adv. Neural Inf. Proces. Syst.* 34, 25956–25967.

Saito, K., and Saenko, K. (2021). *Ovanet: One-vs-all network for universal domain adaptation*. Proceedings of the ieee/cvf international conference on computer vision.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). *Grad-cam: Visual explanations from deep networks via gradient-based localization*. Proceedings of the IEEE international conference on computer vision.

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6, 1–48. doi: 10.1186/s40537-019-0197-0

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., et al. (2020). Fixmatch: simplifying semi-supervised learning with consistency and confidence. *Adv. Neural Inf. Proces. Syst.* 33, 596–608.

Uddin, A. S., Monira, M. S., Shin, W., Chung, T., and Bae, S. H. (2006). *SaliencyMix: a saliency guided data augmentation strategy for better regularization*. International Conference on Learning Representations.

Wang, A., Chen, H., Lin, Z., and Han, J., G. Ding (2024). *Repvit: Revisiting mobile cnn from vit perspective*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Xiaojin, Z. (2008). *Semi-supervised learning literature survey*. Computer Sciences TR, No. 1530.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. (2019). *Cutmix: Regularization strategy to train strong classifiers with localizable features*. Proceedings of the IEEE/CVF international conference on computer vision.

Zagoruyko, S., and Komodakis, N. (2016). *Wide residual networks*. British Machine Vision Conference 2016; British Machine Vision Association.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2018). *Mixup: Beyond empirical risk minimization*. International conference on learning representations.

Zhao, X., Krishnateja, K., Iyer, R., and Chen, F. (2022). *How out-of-distribution data hurts semi-supervised learning*. 2022 IEEE International Conference on Data Mining (ICDM), IEEE.

Zhu, Y., Ma, J., Zhang, Z., Zhang, Y., Zhu, S., Liu, M., et al. (2022). Automatic view classification of contrast and non-contrast echocardiography. *Front. Cardiovasc. Med.* 9:989091. doi: 10.3389/fcvm.2022.989091

Zhu, L., Wang, X., Ke, Z., Zhang, W., and Lau, R. W. (2023). *Biformer: Vision transformer with bi-level routing attention*. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.