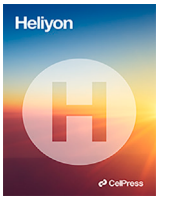
Heliyon 10 (2024) e23224

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

STGA-MS: AI diagnosis model of regional wall motion abnormality based on 2D transthoracic echocardiography

Song Sun [a,b,1], Yonghuai Wang [c,1], Qi Yu [a,b], Mingjun Qu [a,b], Honghe Li [a,b], Jinzhu Yang [a,b,*]

[a] Computer Science and Engineering, Northeastern University, Shenyang, China
[b] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, China
[c] Department of Cardiovascular Ultrasound, The First Hospital of China Medical University, China

ARTICLE INFO

Keywords:
Spatial-temporal grouping attention
Segment-related feature
2D transthoracic echocardiography
RWMA diagnosis

ABSTRACT

Regional wall motion abnormality (RWMA) is a common manifestation of ischemic heart disease detected through echocardiography. Currently, RWMA diagnosis heavily relies on visual assessment by doctors, leading to limitations in experience-based dependence and suboptimal reproducibility among observers. Several RWMA diagnosis models were proposed, while RWMA diagnosis with more refined segments can provide more comprehensive wall motion information to better assist doctors in the diagnosis of ischemic heart disease. In this paper, we proposed the STGA-MS model which consists of three modules, the spatial-temporal grouping attention (STGA) module, the segment feature extraction module, and the multiscale downsampling module, for the diagnosis of RWMA for multiple myocardial segments. The STGA module captures global spatial and temporal information, enhancing the representation of myocardial motion characteristics. The segment feature extraction module focuses on specific segment regions, extracting relevant features. The multiscale downsampling module analyzes myocardial motion deformation across different receptive fields. Experimental results on a 2D transthoracic echocardiography dataset show that the proposed STGA-MS model achieves better performance compared to state-of-the-art models. It holds promise in improving the accuracy and reproducibility of RWMA diagnosis, assisting clinicians in diagnosing ischemic heart disease more reliably.

1. Introduction

Regional wall motion abnormality (RWMA) is common in ischemic heart disease. For patients with emergency chest pain, identification of RWMA using echocardiography is a recommended method (Class I) by the European Society of Cardiology (ESC) guidelines [1]. Moreover, RWMA is an independent predictor of major adverse cardiovascular events (MACE) in coronary heart disease patients [2]. Currently, RWMA is mainly recognized by visual "eyeballing" of endocardium displacement and ventricular wall thickening rate of myocardium in cardiac ultrasound images [1]. This method has the limitation of experience dependence, strong subjectivity, and suboptimal reproducibility between observers [3]. Inexperienced doctors have a higher rate of misdiagnosis [4]. Existing automatic diagnosis methods have some limitations, such as simplifying the myocardial segmentation criteria or manually selecting ultrasound

* Corresponding author at: Computer Science and Engineering, Northeastern University, Shenyang, China.
E-mail address: yangjinzhu@cse.neu.edu.cn (J. Yang).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2023.e23224
Received 29 December 2022; Received in revised form 27 October 2023; Accepted 29 November 2023

Available online 5 December 2023

frames. Therefore, there is an urgent need for an objective and reproducible RWMA diagnosis method designed for fine myocardial segment criteria that overcomes the limitations of the current approach. The proposed solution aims to address these challenges and provide an innovative and reliable method for RWMA diagnosis, reducing the reliance on physician experience, improving diagnostic accuracy, and enhancing consistency among different doctors.

With the development of artificial intelligence (AI), convolutional neural networks (CNN) have made great progress on several medical image processing problems, such as disease diagnosis [5], lesion location [6], and tissue segmentation [7]. For example, Zhu et al. [8] proposed a brain tumor segmentation method that combines deep semantics and edge information from multimodal MRI. Xu et al. [9] introduced a hybrid feature extraction network that integrated CNNs and Transformers to leverage their respective advantages in feature extraction, along with a multi-dimensional statistical feature extraction module to enhance low-dimensional texture features and improve the segmentation performance of medical images. Li et al. [10] proposed the X-Net, a dual encoding-decoding structure that combines CNNs and Transformers for medical image segmentation. It integrates convolutional downsampling and Transformer encoders to extract local and global features and incorporates a variational auto-encoder branch in the decoding phase to mitigate data insufficiency effects. These methods employed well-established deep learning models, including CNNs and Transformers, as feature extractors for medical image segmentation and classification tasks, resulting in notable advancements in accuracy. Strong feature extraction ability, fast response, and stable performance make CNNs become the implementation standard of image segmentation and classification [11].

There were a few studies for RWMA classification based on echocardiography. In the beginning, feature engineering was the main method. Shalbaf et al. [12] proposed a method based on nonlinear dimensionality reduction for RWMA classification. After that, the deep learning method received wide attention because of its powerful feature representation ability and has been applied to RWMA classification. Omar et al. [13] compared the traditional random forest (RF) model with the CNN model for RWMA diagnosis. Huang et al. [14] trained a deep neural network designed for dynamic video frames to accomplish the recognition of left ventricle (LV) RWMA. Kusunose et al. [15] investigated 2D-CNN for RWMA classification which takes three specific frames (end-diastolic, mid-systolic, and end-systolic phases) from a short-axis view as input. The above methods have some limitations. For example, some methods use static frames as input instead of dynamic frames, and the standard of the myocardial segment had been simplified. Though these methods had important clinical value, RWMA classification based on a more refined myocardial segment standard can provide more comprehensive wall motion information, which is better to assist doctors in the diagnosis of ischemic heart disease. Hence, the RWMA classification problem needs to be further studied.

Given that the apical 2-chamber (A2C) view, apical 3-chamber (A3C) view, and apical 4-chamber (A4C) view can provide global segmental information, we chose these three views to obtain more detailed myocardial motion information. The myocardium in each view was divided into six segments by optical flow tracking. However, the finer myocardial segment had higher requirements for model performance. To address the above problem, we proposed STGA-MS which consists of three modules for the diagnosis of RWMA for multiple myocardial segments based on echocardiography. There are three main steps to achieve this goal. First, the SFNet [16] was used to obtain the myocardial segmentation results. Then an optical flow tracking segment method was applied to produce myocardial segments belonging to three different views. Finally, we used the STGA-MS to complete the diagnosis of myocardial segment RWMA.

There are three contributions:

(1) We propose a method called STGA-MS for the objective and reproducible diagnosis of LV RWMA. This method utilizes fine myocardial segment criteria, providing a more robust and reliable approach to RWMA diagnosis.

(2) We proposed the spatial-temporal grouping attention (STGA) module, designed specifically to extract change features of myocardial motion in both the temporal and spatial dimensions. This module enhances the ability to capture and analyze dynamic patterns in myocardial motion, leading to improved accuracy and effectiveness in RWMA diagnosis.

(3) We propose two modules in our method: the segment feature extraction module and the multiscale downsampling module. The segment feature extraction module captures segment-related features, while the multiscale downsampling module handles scale changes during myocardial motion. These modules enhance the accuracy and effectiveness of RWMA diagnosis by considering localized information and multiscale dynamics.

The remainder of this paper is organized as follows: Section 2 introduces the overall framework for RWMA diagnosis and the proposed STGA-MS model. Section 3 presents the data preparation and the experiment results. We present a discussion in section 4. Finally, section 5 presents the conclusions.

2. Methods

Written informed consent was obtained from all individuals prior to enrollment. The study protocol was approved by the China Medical University Ethics Committee. The number is AF-S0P-07-1.1-01. The study was conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki.

2.1. The overall framework for RWMA diagnosis

The overall framework for RWMA diagnosis can be divided into three main parts as depicted in Fig. 1. The first part was to obtain the myocardium segmentation mask. It took echocardiography as input, and the output is a myocardial segmentation mask. The SFNet [16] was used as the myocardial segmentation model because it can learn semantic flow between feature maps of adjacent levels, which is more conducive to extracting semantic information from echocardiography. It is worth noting that SFNet is limited

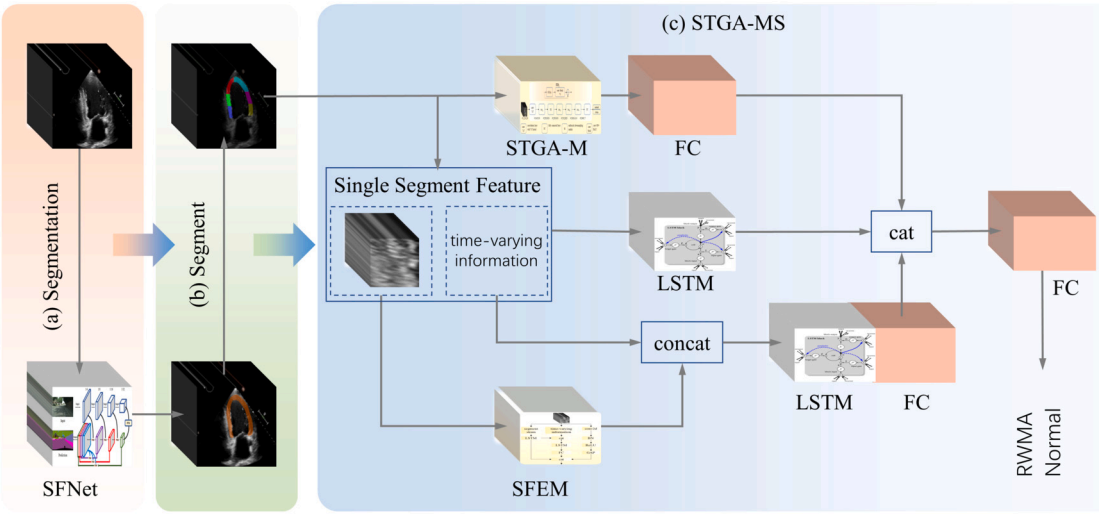


Fig. 1. The overall workflow for LV RWMA diagnosis model. (a) Segmentation: A trained SFNet was used to obtain a myocardial mask for every frame. (b) Segment: Optical flow tracking was used to divide the myocardial into six continuous segments. (c) STGA-MS: The proposed STGA-MS was used to diagnose RWMA of multiple myocardial segments. The SFEM represents the segment feature extraction module. The cat represents concatenate. In the STGA-MS, global feature and segment feature was used to improve the performance. The single segment feature contains two parts, the first one is the echocardiography of a rectangular region surrounding a myocardial segment; the second one is some time-varying information such as the mean pixel value of the myocardial segment in echocardiography. We also applied LSTM [18] to extract the timing information.

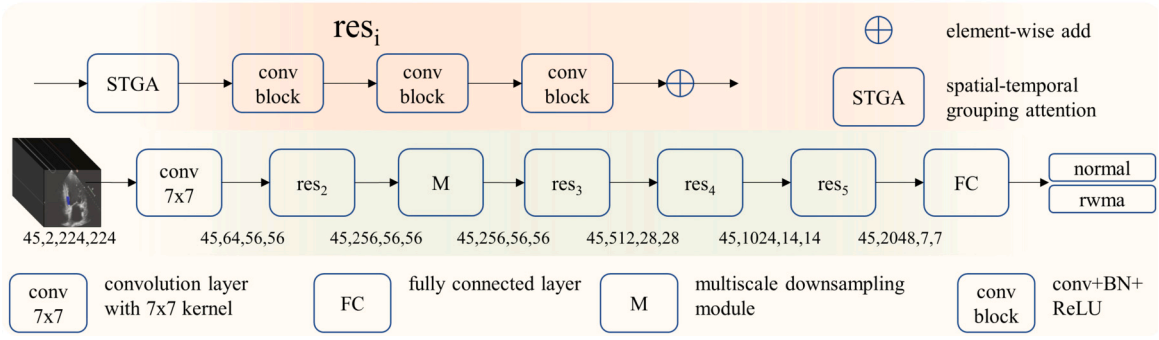


Fig. 2. The structure of resnet18 with STGA and multiscale downsampling module (STGA-MS). The STGA module was placed in front of the three conv blocks in every resblock (res_i). The FC represents a fully connected layer. The conv 7×7 represents the convolution layer with a 7×7 kernel. The M represents the multiscale downsampling module, and the conv block represents convolution + batch normal + ReLU.

to generating the myocardium mask as a whole and cannot produce individual myocardial segments correctly. The reason for utilizing SFNet for myocardial segmentation is that the division of myocardial segments is primarily based on knowledge of cardiac anatomy rather than feature differences in echocardiography. Therefore, SFNet is effective in recognizing the entire myocardial mask. However, it may not accurately divide the myocardial mask into multiple segments. We solve this problem in the second part. The second part was to divide the myocardium segmentation mask into six segments by using optical flow tracking. By this method, we can achieve the myocardial segments with the required quality which are inspected by the doctors. The third part is the RWMA classification model (STGA-MS) using dynamic echocardiography frames and the corresponding myocardial segments as input. The RWMA labels include normal and abnormal, which represent the presence or absence of RWMA for a single segment, respectively. Three different modules were proposed to improve the performance of the RWMA classification, they are STGA, segment feature extraction module, and multiscale downsampling module. Fig. 2 shows the structure of resnet18 [17] with STGA and multiscale downsampling module (STGA-M).

2.2. LV myocardial segmentation

Given that the purpose of segmentation is to acquire a general position for RWMA classification, we employed SFNet [16] for myocardial segmentation without optimizing the model. For training the model, we constructed a dataset consisting of three views: apical two-chamber (A2C), apical three-chamber (A3C), and apical four-chamber (A4C). The dataset was divided into training, validation, and test sets with a ratio of 8:1:1. Further details regarding this dataset can be found in Section 3.1.

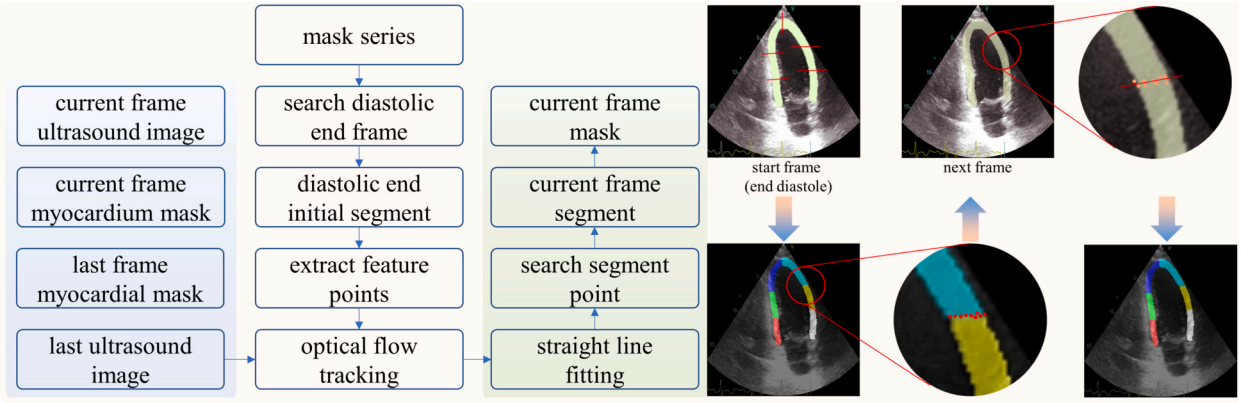


Fig. 3. Flow diagram of dividing the myocardial mask into six segments using the optical flow tracking method. We first find the diastolic end frame by looking for the longest midline of the myocardium. After that, the myocardium in the diastolic end frame was divided into six segments. Then we use the optical flow tracking method to segment all frames in turn.

2.3. LV myocardial segments tracking

To produce the LV myocardial segment mask which meets clinical criteria, we utilized a method based on optical flow tracking to produce 6 segments for each frame. Fig. 3 shows the process step of optical flow tracking. The flow diagram on the left represents the process of generating myocardial segments. On the right, a detailed illustration showcases the steps involved in the myocardial segmentation process. The input consists of a series of myocardial masks and the corresponding echocardiography series, while the output represents the resulting myocardial segments. The Gunnar Farneback algorithm [19] implemented by OpenCV was used for optical flow tracking. Gunnar Farneback algorithm is based on the assumption of image gradient constancy and local optical flow constancy, and it has good stability.

The detailed process is outlined below. Firstly, the end-diastolic frame is determined by traversing and calculating the length of the myocardial mask's center line for all frames. This end-diastolic frame serves as the starting point in the right part of Fig. 3.

Secondly, following the segment standard [20], the myocardial mask of the end-diastolic frame is divided into six segments. In the right part of Fig. 3, you can observe the five red lines that separate the myocardium into six segments.

Next, the feature points are extracted from the dividing lines and utilized for optical flow tracking. These points can be seen at the bottom center of the right part of Fig. 3. By tracking these feature points, the corresponding points in the subsequent frame are identified. To avoid the loss of feature points during optical flow tracking, several points near the separation lines are selected to participate in the tracking process, rather than solely relying on the starting and ending points of a line segment.

Once the feature points of the next frame are tracked, a straight line is fitted based on these points, as depicted in the top right of the right part of Fig. 3. Finally, the intersection of the fitting line with the myocardium yields the myocardial segment point. This process is repeated until tracking is completed for all frames. As a result, six segments are achieved for each frame using this method.

2.4. Main modules of STGA-MS

STGA-MS consists of three modules, which were STGA module, the segment feature extraction module, and the multiscale downsampling module.

2.4.1. STGA module

The STGA module consists of two components, the temporal extraction module (TEM) and the spatial group extraction module (SGEM). Notations used in this section are N (batch size), T (number of frames), C (channels), H (height), W (width).

Timing information plays a crucial role in recognizing myocardial movement patterns. To capture the temporal characteristics of myocardial motion, we developed the Temporal Encoding Module (TEM). The TEM is specifically designed to extract the variation patterns of myocardial motion in the time dimension by modeling the relationship between adjacent frames. As illustrated in Fig. 4, given an input $X \in \mathbb{R}^{N\times T\times C\times H\times W}$, we used a 2D convolution layer L_1 with kernel size 1×1 to process X as

$$F = L_1 * X \tag{1}$$

Then, a 2D convolution layer L_2 with kernel size 3×3 was used to process $F \in \mathbb{R}^{N\times T\times C\times H\times W}$, and the feature subtraction operation was implemented, which can be represented as

$$F_m^t = L_2 * (F[:, t+1, :, :, :] - F[:, t, :, :, :]) \tag{2}$$

where $F_m^t \in \mathbb{R}^{N\times 1\times C\times H\times W}, t \in [1,\ T-1]$, and there were $T-1$ F_m features. These features were then concatenated with each other according to the temporal dimension, and the last F_m^T was padded with a zero matrix of the same shape. The concatenated feature

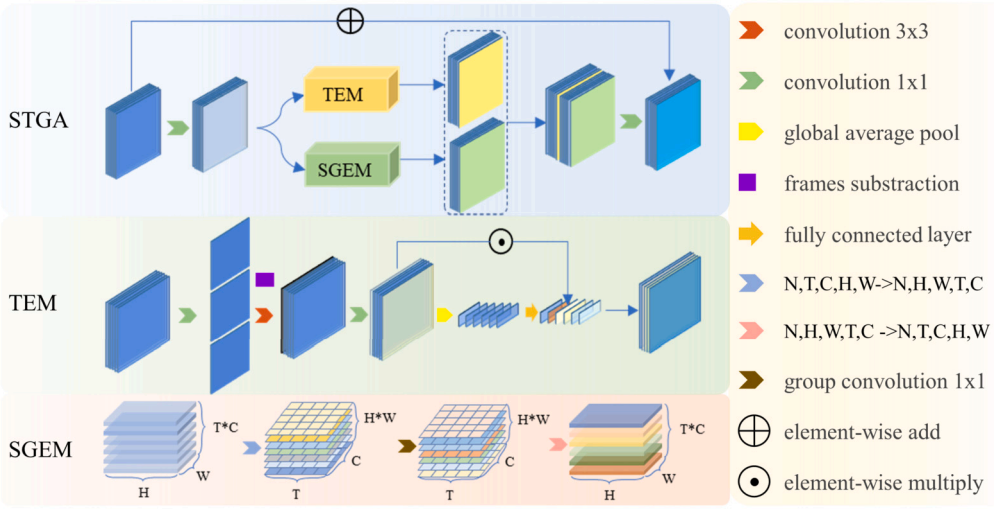


Fig. 4. The architecture of the STGA module. STGA represents spatial-temporal grouping attention. The STGA module consists of the temporal extraction module (TEM) and the spatial group extraction module (SGEM). The TEM and the SGEM were parallel to extract different features.

$F_M = \left[F_m^1, \cdots, F_m^{T-1}, 0\right], F_M \in \mathbb{R}^{N\times T\times C\times H\times W}$. The F_M is then processed with a 2D convolution layer L_3 with kernel size 1×1, which can be represented as

$$F_M^* = L_3 * F_M \tag{3}$$

where $F_M^* \in \mathbb{R}^{N\times 1\times C\times H\times W}$. Then we used a global max pooling layer GMP_1 and a fully connected layer FC_1 to further enhance the feature representation capability. First, F_M^* was reshaped to $F_M^{r1}\in\mathbb{R}^{(N*C)\times T\times H\times W}$, then we used GMP_1 to process F_M^{r1}, the result $F_M^G\in\mathbb{R}^{N\times C\times T}$ was then processed by FC_1. After that, the F_M^* and the result of $\mathrm{FC}_1(F_M^G)$ were implemented pixel-by-pixel multiplication, the output of TEM was $F_{TEM}\in\mathbb{R}^{N\times T\times C\times H\times W}$.

$$F_M^G = \mathrm{GMP}_1\left(F_M^{r1}\right) \tag{4}$$

$$F_{TEM}^{=} F_M^* * \left(\mathrm{FC}_1\left(F_M^G\right)\right) \tag{5}$$

Spatial information can provide high-level features of each frame image, thus improving the stability of the model. SGEM was designed for extracting the feature of the myocardial segment in the spatial dimension. It used a 2D convolution layer L_4^G with kernel size 3×3, and the number of input channels and output channels was both $H\times W$. The group number of L_4 was set to W. In this way, each group just concentrated features from a single row of the image, which is useful for extracting the subtle features of the myocardial segments.

$$X_{norm} = \frac{X}{W}\Big| X\in\{left,\ right\} \tag{6}$$

The STGA consisted of two 2D convolution layers with kernel size 1×1, which were located at the front and back of TEM and SGEM. The TEM was parallel with SGEM. The outputs were then concatenated. In the end, a skip connect was conducted.

2.4.2. Segment feature extraction module

Although CNN can theoretically extract any type of feature, due to the small data scale of medical image data sets, providing specific features related to segment regions can effectively improve model performance. To locate the myocardial segment and extract features that vary with the motion of the myocardial segment, a minimal bounding rectangle was used to delineate the region where the myocardial segment was located. We called this minimal bounding rectangle the neighborhood region, and we called the irregular region where the myocardial segment the segment region.

Fig. 5 depicts the segment feature extraction module. It contained two branches. The left one used some statistical information on the myocardial segment. The statistical information includes the relative coordinates of the upper left and lower right corner of the neighborhood region, the ratio of the area of the segment region to the area of the neighborhood area (*area_ratio_box*), the ratio of the area of the segment region to the area of the whole region (*area_ratio_total*), and the maximum, mean, and minimum pixel values of the segment region of the echocardiography. All of the above values are between 0 and 1. The above statistical information is all directly associated with the motion of the myocardial segment. For example, the *area_ratio_box* is related to the thickness of the myocardial wall. At the end of diastole, the thickness of the myocardial wall reached the minimum and the *area_ratio_box* reaches a minimum as well. At the end of the systole, the myocardial wall thickness reaches a maximum and the *area_ratio_box* reaches a maximum at the same time. The parameters of LSTM are described: input size is 45, hidden size is 1, and layer number is 3.

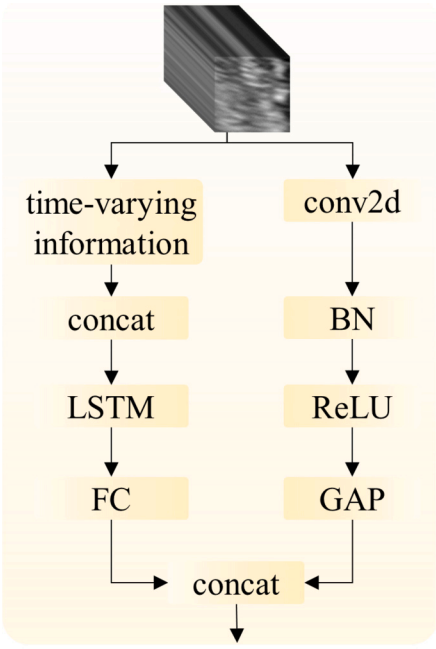


Fig. 5. Segment feature extraction module. It contains two branches. The first branch extracts time-varying information based on LSTM. The second branch extracts high-level features based on convolution.

The right one used the neighborhood region of the echocardiography as input. It first resizes the input to 64×64. Then 2D convolution, batch normalization, ReLU, 2D convolution, and the global average pool was used to extract high-level segment feature. Finally, the features from the three branches were concatenated. The relevant formula is given in the next paragraph. The conv2d shown in Fig. 5 is two 2D convolution layers comprised of 64 3×3 kernels, and the stride used is 2.

$$X_{norm} = \frac{X}{W} \Big| X \in \{left,\ right\} \tag{7}$$

$$Y_{norm} = \frac{Y}{H} \Big| Y \in \{up,\ down\} \tag{8}$$

$$\left.\begin{array}{l} max\_value = MAX\left(I_{i,j}\right) \\ min\_value = MIN\left(I_{i,j}\right) \\ mean\_value = MEAN\left(I_{i,j}\right) \end{array}\right| i \in [up,\ down], j \in [left,\ right] \tag{9}$$

$$area\_ratio\_total = \left(\sum_{j=0}^{W-1}\sum_{i=0}^{H-1} M_{i,j}\right) / (H \times W) \tag{10}$$

$$area\_ratio\_box = \left(\sum_{j=0}^{W-1}\sum_{i=0}^{H-1} M_{i,j}\right) / ((right - left + 1) \times (down - top + 1)) \tag{11}$$

2.4.3. Multiscale downsampling module

During myocardial motion, the size of the myocardium undergoes dynamic changes, making it challenging to extract consistent and effective features from fixed scales alone. Therefore, it is crucial to extract myocardial features at multiple scales in order to capture comprehensive information. As depicted in Fig. 6, the input was processed by three branches. Each branch contained several downsampling 2D convolutions with stride 2×2, padding 1×1, and kernel 3×3. Each of the downsampling 2D convolution layers reduced the size of the feature map by half. The first branch contains three downsampling 2D convolution layers. The second branch contains two downsampling 2D convolution layers, and the input is the output of the first convolution layer of the first branch. The third branch contains one downsampling 2D convolution layer, and the input is the output of the first convolution of the second branch. To extract richer deep features, multiple data-sharing operations were conducted. Finally, the outputs of the three branches were concatenated. The numbers of output channels were 128, 128, and 64, which were marked on the feature map of each branch. The final output feature had the shape of $(H/8, W/8, 320)$. H and W were the height and width of the input feature map. The idea of the multiscale downsampling module came from HRNet [21], the difference is that our module did not contain additional processes such as resample when concatenating features from a different branch.

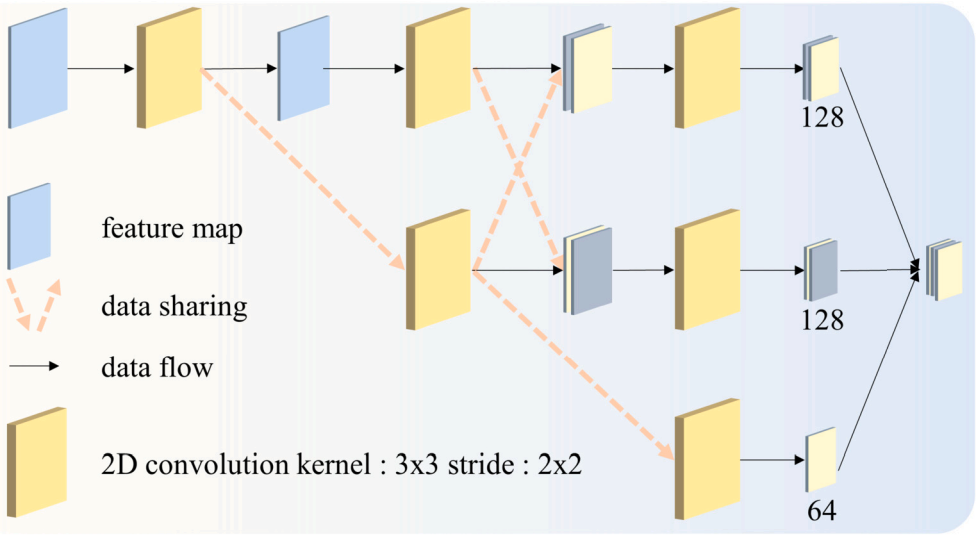


Fig. 6. Multiscale downsampling module. This module performs multi-scale feature aggregation through three parallel branches. The output feature size is 1/8 of the input.

Table 1
The detail of the myocardial segmentation dataset. Every patient has A2C, A3C, and A4C view.

	Patients	A2C (frames)	A3C (frames)	A4C (frames)	Total (frames)
Train set	72	2794	2610	2637	8041
Validation set	9	327	383	376	1086
Test set	10	221	300	438	959
Total	91	3342	3293	3451	10086

3. Results

3.1. Dataset description and implementation details

In this study, we curated two distinct datasets for our research purposes. The first dataset was utilized to train the myocardial segmentation model, while the second dataset was employed for the RWMA classification model. Importantly, there was no overlap between the two datasets to ensure independent evaluation. Both datasets were meticulously collected and labeled by two experienced doctors, each possessing five years of clinical experience, using the labelme tool. Subsequently, all the labels were thoroughly reviewed and examined by two doctors with a decade of clinical experience. In cases where there were discrepancies in the data assessment between the two experts, the final determination was made by a third doctor, also with ten years of clinical experience. These medical professionals are associated with The First Hospital of China Medical University.

The myocardial segmentation dataset was all from GE equipment. In this dataset, 8041 frames of 72 patients as the training dataset, 1086 frames of 9 patients as the validation dataset, and 959 frames of 10 patients as the test dataset. Data belonging to the same patient are grouped into the same set. Each patient contains three views, A2C, A3C, and A4C. The detail of this dataset is presented in Table 1.

All experiments were implemented on GPU NVIDIA Tesla T4 with Intel(R) Xeon(R) Gold 5215 CPU @ 2.50 GHz, Pytorch 1.10.2. For all the experiments, the data were normalized to accelerate convergence. The learning rate was set to 0.001, and Adam was used as the optimizer. The batch size was set to 5 for the segmentation model and 6 for the RWMA classification model. The number of epochs for all experiments was 50.

The RWMA classification dataset with A2C, A3C, and A4C views from 148 patients was constructed. All data were produced by the GE device. Due to the poor quality of a few echocardiogram frames, we removed those with poor myocardial segmentation results using an automatic filtering method. The automatic filtering method removes frames with poor myocardial segmentation results, including broken myocardial segmentation results and severe under-segmentation results. Finally, we constructed an RWMA dataset with 137 patients. The filtered data accounted for 7.43% of the total dataset. Among these data, 82 patients had no RWMA in all three views, and 55 patients had RWMA. We split the data using five-fold cross-validation as depicted in Table 2. Patients with and without RWMA were stratified by random sampling. In addition, there were 2057 normal segments and 409 segments with RWMA, and the ratio of it is 5:1.

Table 2
The detail of the myocardial segmentation dataset. There are patients with RWMA and without RWMA in this dataset.

	Patients	without RWMA	with RWMA
train	109	65	44
test	28	17	11
total	137	82	55

Table 3
Ablation experiments for multiscale downsampling module with 1 to 3 branches.

	AUC		specificity		sensitivity		accuracy	
	mean	std	mean	std	mean	std	mean	std
3-branch	**0.6730**	0.0633	0.8052	0.0898	**0.4845**	0.1825	**0.7622**	0.0630
2-branch	0.6219	0.1136	0.8386	0.0649	0.3273	0.1980	0.7612	0.0377
1-branch	0.6343	0.0364	**0.8475**	0.0693	0.2552	0.0800	0.7519	0.0606

3.2. LV myocardial segmentation results

The Dice function in Eq. (12) was used to evaluate the segmentation experiment results, where P was the predicted result and T was the ground truth.

$$Dice = \frac{2|P \cap T|}{|P| + |T|} \tag{12}$$

The Dice of A2C is 0.8740, A3C is 0.8323, and A4C is 0.8719. The results show that the model can accurately segment myocardium and can be used for RWMA classification.

3.3. LV myocardial segment tracking result

This part is implemented following clinical standards without ground truth, only visualized images of myocardial segmentation superimposed with echocardiography were presented. As depicted in Fig. 7, every two rows represented a view, which was the echocardiography and corresponding myocardial segment mask respectively, and every column represented a patient. Fig. 7 shows that the myocardial segment is accurately tracked and this method works for three views.

3.4. Ablation experiments

To determine the optimal number of branches for the multiscale downsampling module, we performed experiments ranging from 1 to 3 branches. We did not include a 4-branch experiment because downsampling to 1/16 resolution would not effectively capture spatial features. The corresponding results are shown in Table 3. Among the tested configurations, the multiscale downsampling module with 3 branches (referred to as the 3-branch module) demonstrated the highest AUC, sensitivity, and accuracy. Based on these findings, we decided to set the branch number of the multiscale downsampling module to 3 for subsequent analyses.

In order to thoroughly investigate the individual contributions of each proposed module on model performance, an ablation experiment was conducted. In this experiment, we designated the model incorporating the STGA and segment feature extraction module as STGA-S and the STGA-S model augmented with the multiscale downsampling module as STGA-MS. This systematic approach allowed us to assess the specific impact of each module on the overall performance of the model.

As presented in Table 4, the STGA-MS model demonstrated the best performance in terms of specificity and accuracy. When comparing it to the STGA-S model, which lacks the multiscale downsampling module, a slight reduction in mean specificity and mean accuracy was observed, while the mean AUC and mean sensitivity increased. We attribute this to the fact that the multiscale downsampling module focuses on capturing overall multiscale information. However, since the scale variation of myocardial segments is relatively small compared to the size of the echocardiography, the impact of the multiscale downsampling module is somewhat limited.

Furthermore, when the segment feature module was removed, the STGA model achieved a mean AUC of 0.6708, a mean specificity of 0.7855, a mean sensitivity of 0.4566, and a mean accuracy of 0.7299. In comparison, the resnet18 model [17] achieved a mean AUC of 0.6459 and a mean specificity of 0.6821. On the other hand, the STGA-MS model achieved a mean AUC of 0.6730 and a mean specificity of 0.8052. These results indicate that the STGA-MS model aligns well with the characteristics of RWMA diagnosis tasks, further reinforcing its suitability for such tasks.

The above ablation experiments show that the proposed modules all promote the performance improvement of the model. Fig. 8 shows the five-fold ROC curves of STGA, STGA-S, and STGA-MS. The results of the second fold are generally lower than other folds, which shows that the second fold is more challenging. This indicates that the dataset we used possesses data of different quality and is closer to the clinical scenario.

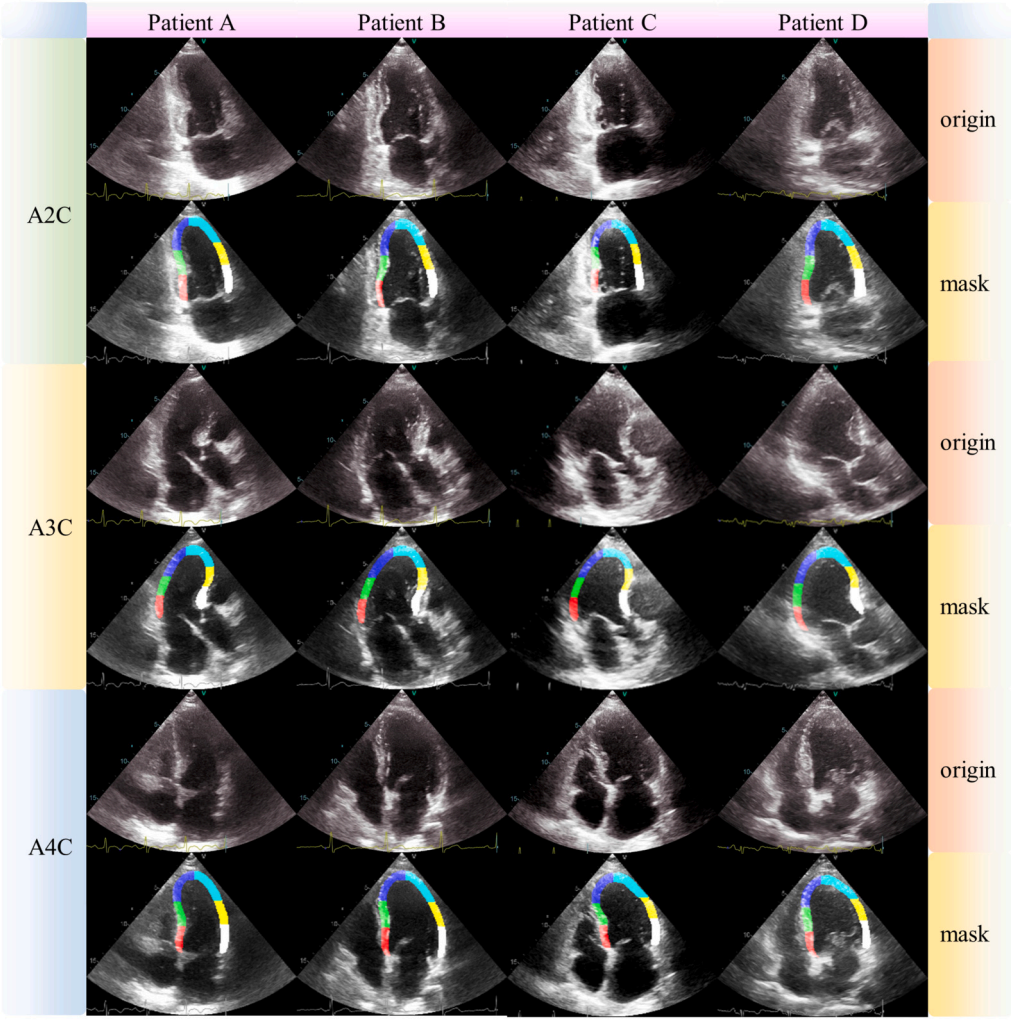


Fig. 7. The visual result for dividing the myocardial mask. We choose four typical patients named patientA to patientD. Three views (A2C, A3C, A4C) for every patient were present.

Table 4
Ablation experiments result. We show the impact of modules on model performance by sequentially removing modules.

Method	AUC		specificity		sensitivity		accuracy	
	mean	std	mean	std	mean	std	mean	std
STGA-MS	0.6730	0.0633	**0.8052**	0.0898	0.4845	0.1825	**0.7622**	0.0630
STGA-S	**0.6982**	0.1243	0.7533	0.0189	0.5324	0.1922	0.7240	0.0301
STGA	0.6708	0.0660	0.7855	0.1123	0.4566	0.1993	0.7298	0.0699
resnet18 [17]	0.6459	0.0790	0.6821	0.0353	**0.5576**	0.0939	0.6658	0.0332

The mean AUC of STGA-S is 0.0275 higher than STGA, while the STGA-S had a higher variance. This indicates that segmental features can improve the performance of the model, but reduce its stability. When adding the multiscale downsampling module, the mean AUC of STGA-MS reduced to 0.6272, while the variance is 0.0633, which is the lowest. It reveals that the multiscale downsampling module can effectively improve the stability of the model.

Furthermore, we present the loss plot of STGA, STGA-S, and STGA-MS in the training process in Fig. 9. It is clear that the loss value decreases with the increase of the training epoch, and the loss value fluctuates to a certain extent during the training process and finally reaches stability.

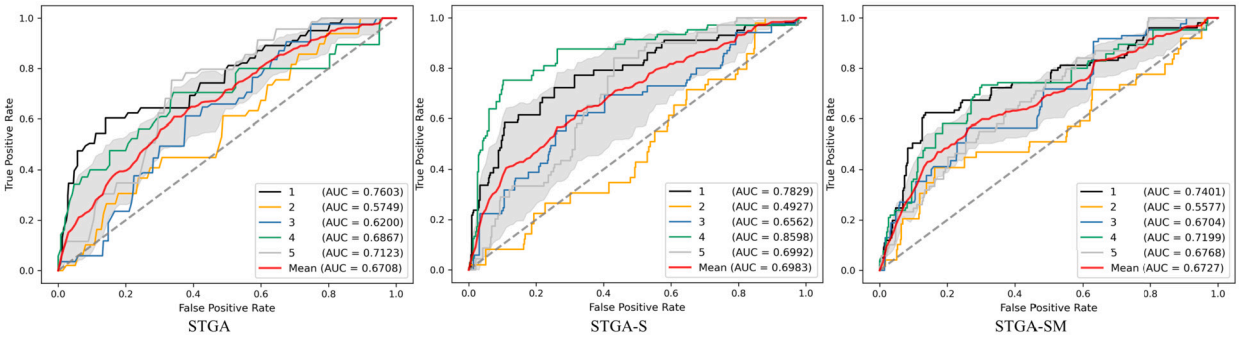


Fig. 8. The overall ROC curve for STGA, STGA-S, STGA-MS. Numbers 1 to 5 in the legend represent five-fold cross-validation. Mean represents the mean AUC of five-fold cross-validation. The gray area represents the variance range. The red line represents the average receiver operator characteristic curve (ROC) of five-fold cross-validation.

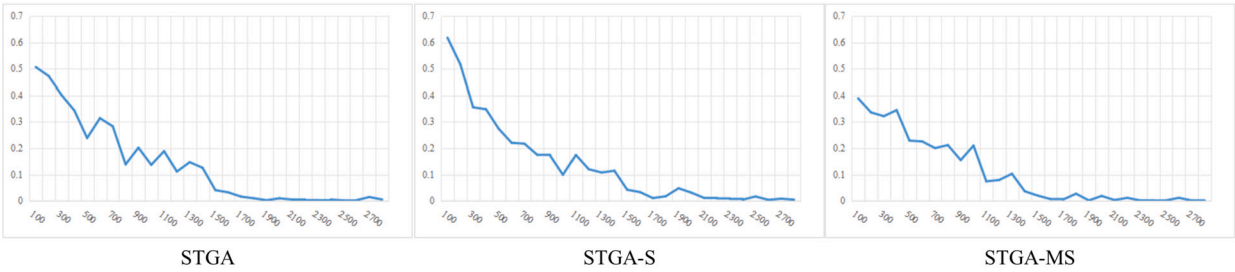


Fig. 9. The loss plot in the training process of STGA, STGA-S, and STGA-MS from left to right. The horizontal axis represents the training steps. The value of training steps is not the epoch, since we just save the loss value every 100 backpropagations. The vertical axis represents the loss value.

Table 5
Performance comparison of the proposed model STGA-MS against other models on the five-fold cross-validation. "-" in this table represent this model can not be trained for RWMA classification effectively.

Method	AUC		specificity		sensitivity		accuracy	
	mean	std	mean	std	mean	std	mean	std
C3D [22]	-	-	-	-	-	-	-	-
TSN [23]	-	-	-	-	-	-	-	-
TSM [24]	-	-	-	-	-	-	-	-
TAM [25]	0.6287	0.0484	0.6094	0.1388	**0.5273**	0.1470	0.5894	0.1074
ACTION [27]	0.6109	0.0519	0.7179	0.2183	0.3872	0.2257	0.6619	0.1368
AIA [26]	0.6063	0.0445	0.5997	0.1608	0.4899	0.2006	0.5734	0.0908
Huang2020 [14]	0.5483	0.0892	0.7351	0.0344	0.3254	0.1040	0.6664	0.0360
Kusunose2020 [15]	0.5952	0.0766	0.7989	0.0306	0.3076	0.1216	0.7188	0.0390
STGA-MS	**0.6730**	0.0633	**0.8052**	0.0898	0.4845	0.1825	**0.7622**	0.0630

3.5. Comparison of RWMA classification results

We compare our approach with the state-of-the-art video classification methods on the RWMA classification task. Table 5 shows the comparison result between these methods. We compared our STGA-MS with C3D [22], TSN [23], TSM [24], TAM [25], AIA [26], Huang2020 [14], and Kusunose2020 [15]. The C3D [22], TSN [23], and TSM [24] can not be effectively trained for this dataset. Hence, we used "-" instead of numerical values in the corresponding positions in this table. Our STGA-MS achieves the best results compared with other methods in terms of AUC, specificity, and accuracy. Although the TAM [25] achieved a higher sensitivity than STGA-MS, the other three metrics are lower than STGA-MS. Specifically, the specificity of STGA-MS is 0.1958 higher than TAM [25]. And the standard deviation of STGA-MS is relatively smaller. It revealed the STGA-MS has better stability. Moreover, Huang2020 [14], and Kusunose2020 [15] are two methods for RWMA classification. However, it is important to note that these methods are designed for simplified myocardial segment criteria, which results in lower performance compared to our proposed method, STGA-MS. The main reason is that STGA-MS consisted of three modules designed for effectively and efficiently extracting the motion and deformation information of the myocardial segments.

4. Discussion

4.1. Difference between STGA-MS and state-of-the-art models

As depicted in Table 4, the STGA-MS achieves a mean AUC of 0.6730, while C3D [22], TSN [23], and TSM [24] cannot be effectively trained. The main reason is that the STGA-MS is specifically designed for myocardial motion characteristics, while other models just focused on natural scene video classification. The image quality of natural scene video is better than that of echocardiography, and the motion of natural scene object is simpler than that of the myocardium. Hence, some of the above models can not be effectively trained, and others had a lower performance than STGA-MS, STGA-MS achieved the best result.

In the STGA module, there are two sub-modules. In the TEM sub-module, the pooling layer and full connection layer were applied to obtain the weight of the temporal dimension and to achieve the effect of inter-frame modeling based on dynamic information, which is different from the existing methods. In the SGEM sub-module, a 2D convolution layer was used to extract features in the temporal dimension and channel dimension synchronously. In addition, to make the SGEM more concentrated on modeling the relationship between the temporal dimension and channel dimension, some 2D convolution layers that were consistent with the number of rows in the feature map were used to extract features from each row and then integrate all features in the row dimension. In this way, the STGA module can pay more attention to local motion information.

In the segment feature extraction module, we consider two aspects of information. One is the echocardiography of the segment region, the other is the information from the segment mask. The echocardiography of the segment region is used to assist the model to focus on the segment region, while the segment mask provides some statistical information such as relative coordinates and area ratio.

The multiscale downsampling module is mainly used to extract the global multiscale features of the echocardiography. In addition, the segment longitudinal strain was used in the model since it is used as a diagnostic basis in clinical practice.

4.2. Comparison with traditional methods

There are three recent studies about RWMA classification. Kusunose et al. [15] used 3 mid-level short-axis static images belonging to end-diastolic, mid-systolic, and end-systolic phases as model input to detect the presence of RWMA based on the resnet50 model. The use of manually selected specific frames as input is not consistent with the clinical scenario using dynamic video. Huang et al. [14] used a DenseNet-based [28] model to diagnose RWMA using dynamic videos. The myocardium from four short-axis views was only divided into four categories, with some simplification based on the 2015 American Society of Echocardiography guidelines. Lin et al. [29] used R2plus1D [30] for detecting RWMA. The myocardium from three views was divided into three categories in sum. In addition, this model is also a single-label model.

There are two limitations to our paper. Firstly, the STGA-MS model employed in this study is a single-label classification model, which assigns a single label to each view, instead of a multi-label classification model that would allow for multiple labels to be assigned to a single view. The decision to use a single-label model was driven by the constraints of the RWMA dataset used in this study, which did not provide sufficient support for the task of multi-label classification. Secondly, our STGA-MS model focuses solely on detecting the presence of myocardial segment RWMA and does not provide a comprehensive four-grade scoring system that includes normal, hypokinetic, akinetic, and dyskinetic segments. Wall motion score indexes are commonly used for diagnosing ischemic heart disease, and a more detailed scoring system would provide valuable insights. However, due to the increasing severity of RWMA, the occurrence of akinetic and dyskinetic segments becomes less frequent, leading to a severe class imbalance problem. This imbalance poses challenges in automatically assigning wall motion scores. Additionally, the scarcity of available RWMA datasets presents a challenge in obtaining external validation sets. To address this limitation, our future research will focus on generating more extensive patient cohorts through collaboration with multiple medical centers. This collaborative effort aims to develop an automated four-grade scoring model for regional wall motion abnormality.

5. Conclusions

In this paper, we proposed an LV RWMA diagnosis model named STGA-MS, which can diagnose the presence of RWMA for multiple myocardial segments belonging to three different views based on echocardiography. Three modules named STGA, the segment feature extraction module, and the multiscale downsampling module were proposed based on myocardial motion characteristics. In addition, myocardial segment longitudinal strain was applied to the proposed model. Compared with the state-of-the-art models, the STGA-MS achieved a mean AUC of 0.6730 in the five-fold cross-validation. The ablation studies show that every module proposed in STGA-MS contributes to the model performance.

CRediT authorship contribution statement

Song Sun: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Yonghuai Wang:** Data curation. **Qi Yu:** Data curation. **Mingjun Qu:** Formal analysis. **Honghe Li:** Investigation. **Jinzhu Yang:** Supervision, Resources, Conceptualization.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. U21A20387, U22A2022)), the Liaoning Province "Xingliao Talents Plan" project (No. XLYC1905001), the Project from Department of Science & Technology of Liaoning Province (No. 2021JH1/10400051).

References

[1] M. Roffi, C. Patrono, J.P. Collet, C. Mueller, M. Valgimigli, F. Andreotti, J.J. Bax, M.A. Borger, C. Brotons, D.P. Chew, B. Gencer, G. Hasenfuss, K. Kjeldsen, P. Lancellotti, U. Landmesser, J. Mehilli, D. Mukherjee, R.F. Storey, S. Windecker, H. Baumgartner, O. Gaemperli, S. Achenbach, S. Agewall, L. Badimon, C. Baigent, H. Bueno, R. Bugiardini, S. Carerj, F. Casselman, T. Cuisset, C. Erol, D. Fitzsimons, M. Halle, C. Hamm, D. Hildick-Smith, K. Huber, E. Iliodromitis, S. James, B.S. Lewis, G.Y.H. Lip, M.F. Piepoli, D. Richter, T. Rosemann, U. Sechtem, P.G. Steg, C. Vrints, J.L. Zamorano, 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC), Eur. Heart J. 37 (3) (2016) 267–315, https://doi.org/10.1093/eurheartj/ehv320.

[2] F. Rigo, R. Sicari, S. Gherardi, A. Djordjevic-Dikic, L. Cortigiani, E. Picano, The additive prognostic value of wall motion abnormalities and coronary flow reserve during dipyridamole stress echo, Eur. Heart J. 29 (1) (2008) 79–88, https://doi.org/10.1093/eurheartj/ehm527.

[3] A.F. Parisi, P.F. Moynihan, E.D. Folland, C.L. Feldman, Quantitative detection of regional left ventricular contraction abnormalities by two-dimensional echocardiography. II. Accuracy in coronary artery disease, Circulation 63 (4) (1981) 761–767, https://doi.org/10.1161/01.CIR.63.4.761.

[4] Maleeha Qazi, Glenn Fung, Sriram Krishnan, Romer Rosales, Harald Steck, R. Bharat Rao, Don Poldermans, Dhanalakshmi Chandrasekaran, Automated heart wall motion abnormality detection from ultrasound images using Bayesian networks, in: IJCAI, vol. 7, 2007, pp. 519–525.

[5] Sui Peng, Yihao Liu, Weiming Lv, Longzhong Liu, Qian Zhou, Hong Yang, Jie Ren, Guangjian Liu, Xiaodong Wang, Xuehua Zhang, Qiang Du, Fangxing Nie, Gao Huang, Yuchen Guo, Jie Li, Jinyu Liang, Hangtong Hu, Han Xiao, Zelong Liu, Fenghua Lai, Qiuyi Zheng, Haibo Wang, Yanbing Li, Erik K. Alexander, Wei Wang, Haipeng Xiao, Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study, Lancet Digit. Health 3 (4) (2021) e250–e259, https://doi.org/10.1016/S2589-7500(21)00041-8.

[6] Z. Yang, L. Zhao, S. Wu, C.Y. Chen, Lung lesion localization of Covid-19 from chest CT image: a novel weakly supervised learning method, IEEE J. Biomed. Health Inform. 25 (6) (2021) 1864–1872, https://doi.org/10.1109/JBHI.2021.3067465.

[7] Zhijie Zhang, Huazhu Fu, Hang Dai, Jianbing Shen, Yanwei Pang, Ling Shao, ET-Net: a generic edge-attention guidance network for medical image segmentation, in: Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, Ali Khan (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Springer International Publishing, ISBN 978-3-030-32239-7, 2019, pp. 442–450.

[8] Zhiqin Zhu, Xianyu He, Guanqiu Qi, Yuanyuan Li, Baisen Cong, Yu Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, Inf. Fusion 91 (2023) 376–387, https://doi.org/10.1016/j.inffus.2022.10.022.

[9] Yang Xu, Xianyu He, Guofeng Xu, Guanqiu Qi, Kun Yu Li Yin, Pan Yang, Yuehui Yin, Hao Chen, A medical image segmentation method based on multi-dimensional statistical features, Front. Neurosci. 16 (2022) 1009581, https://doi.org/10.3389/fnins.2022.1009581.

[10] Yuanyuan Li, Ziyu Wang Li Yin, Zhiqin Zhu, Guanqiu Qi, Yu Liu, X-Net: a dual encoding–decoding method in medical image segmentation, Vis. Comput. (2021) 1–11, https://doi.org/10.1007/s00371-021-02328-7.

[11] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207, https://doi.org/10.1016/j.media.2019.01.012.

[12] A. Shalbaf, H. Behnam, Z. Alizade-Sani, M. Shojaifard, Automatic assessment of regional and global wall motion abnormalities in echocardiography images by nonlinear dimensionality reduction, Med. Phys. 40 (5) (2013) 052904, https://doi.org/10.1118/1.4799840.

[13] H.A. Omar, A. Patra, J.S. Domingos, P. Leeson, A.J. Noblel, Automated myocardial wall motion classification using handcrafted features vs a deep CNN-based mapping, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2018 (2018) 3140–3143, https://doi.org/10.1109/EMBC.2018.8513063.

[14] Mu-Shiang Huang, Chi-Shiang Wang, Jung-Hsien Chiang, Ping-Yen Liu, Wei-Chuan Tsai, Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography, Circulation 142 (16) (2020) 1510–1520, https://doi.org/10.1161/circulationaha.120.047530.

[15] K. Kusunose, T. Abe, A. Haga, D. Fukuda, H. Yamada, M. Harada, M. Sata, A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images, JACC Cardiovasc. Imag. 13 (2) (2020) 374–381, https://doi.org/10.1016/j.jcmg.2019.02.024, ISSN 1876-7591 (Electronic) 1876-7591 (Linking).

[16] Xiangtai Li, Ansheng You, Zhen Zhu, Houlong Zhao, Maoke Yang, Kuiyuan Yang, Shaohua Tan, Yunhai Tong, Semantic flow for fast and accurate scene parsing, in: Andrea Vedaldi, Horst Bischof, Thomas Brox, Jan-Michael Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, ISBN 978-3-030-58452-8, 2020, pp. 775–793.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[18] Klaus Greff, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber, LSTM: a search space odyssey, IEEE Trans. Neural Netw. Learn. Syst. 28 (10) (2016) 2222–2232, https://doi.org/10.1109/tnnls.2016.2582924.

[19] Gunnar Farnebäck, Two-frame motion estimation based on polynomial expansion, in: Scandinavian Conference on Image Analysis, Springer, 2003, pp. 363–370.

[20] J.U. Voigt, G. Pedrizzetti, P. Lysyansky, T.H. Marwick, H. Houle, R. Baumann, S. Pedri, Y. Ito, Y. Abe, S. Metz, J.H. Song, J. Hamilton, P.P. Sengupta, T.J. Kolias, J. d'Hooge, G.P. Aurigemma, J.D. Thomas, L.P. Badano, Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the eacvi/ase/industry task force to standardize deformation imaging, Eur. Heart J. Cardiovasc. Imaging 16 (1) (2015) 1–11, https://doi.org/10.1093/ehjci/jeu184.

[21] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Deep high-resolution representation learning for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2020) 3349–3364, https://doi.org/10.1109/tpami.2020.2983686.

[22] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri, Learning spatiotemporal features with 3d convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4489–4497.
[23] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, Luc Van Gool, Temporal segment networks: towards good practices for deep action recognition, in: European Conference on Computer Vision, Springer, 2016, pp. 20–36.
[24] Ji Lin, Chuang Gan, Song Han, TSM: temporal shift module for efficient video understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7083–7093.
[25] Zhaoyang Liu, Limin Wang, Wayne Wu, Chen Qian, Tong Lu, TAM: temporal adaptive module for video recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 13708–13718.
[26] Yanbin Hao, Shuo Wang, Pei Cao, Xinjian Gao, Tong Xu, Jinmeng Wu, Xiangnan He, Attention in attention: modeling context correlation for efficient video classification, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 7120–7132, https://doi.org/10.1109/tcsvt.2022.3169842.
[27] Zhengwei Wang, Qi She, Aljosa Smolic, ACTION-Net: multipath excitation for action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13214–13223.
[28] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[29] Xixiang Lin, Feifei Yang, Yixin Chen, Xiaotian Chen, Wenjun Wang, Q. Wang, L. Zhang, H. Guo, B. Liu, L. Yu, Echocardiography-based AI detection of regional wall motion abnormalities and quantification of cardiac function in myocardial infarction, Front. Cardiovasc. Med. 9 (2022), https://doi.org/10.3389/fcvm.2022.903660.
[30] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, Manohar Paluri, A closer look at spatiotemporal convolutions for action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6450–6459.