

MReg: A Novel Regression Model with MoE-based Video Feature Mining for Mitral Regurgitation Diagnosis

Zhe Liu^{1,2*}, Yuhao Huang^{1,2*}, Lian Liu^{1,2,3*}, Chengrui Zhang⁴, Haotian Lin³,
Tong Han^{1,2}, Zhiyuan Zhu^{1,2}, Yanlin Chen^{1,2}, Ruiyue Chen⁵, Dong Ni²,
Zhongshan Gou⁶✉, and Xin Yang^{1,2}✉

¹National-Regional Key Technology Engineering Laboratory for Medical Ultrasound,
School of Biomedical Engineering, Medical School, Shenzhen University, China

²Medical Ultrasound Image Computing (MUSIC) Lab, School of Biomedical
Engineering, Medical School, Shenzhen University, China

³Shenzhen RayShape Medical Technology Co., Ltd, China

⁴School of Physics and Astronomy, The University of Edinburgh, United Kingdom

⁵School of Future Technology, South China University of Technology, China

⁶Center for Cardiovascular Disease, The Affiliated Suzhou Hospital of Nanjing
Medical University, China

Abstract. Color Doppler echocardiography is a crucial tool for diagnosing mitral regurgitation (MR). Recent studies have explored intelligent methods for MR diagnosis to minimize user dependence and improve accuracy. However, these approaches often fail to align with clinical workflow and may lead to suboptimal accuracy and interpretability. In this study, we introduce an automated MR diagnosis model (*MReg*) developed on the 4-chamber cardiac color Doppler echocardiography video (A4C-CDV). It follows comprehensive feature mining strategies to detect MR and assess its severity, considering clinical realities. Our contribution is threefold. First, we formulate the MR diagnosis as a regression task to capture the continuity and ordinal relationships between categories. Second, we design a feature selection and amplification mechanism to imitate the sonographer’s diagnostic logic for accurate MR grading. Third, inspired by the Mixture-of-Experts concept, we introduce a feature summary module to extract the category-level features, enhancing the representational capacity for more accurate grading. We trained and evaluated our proposed *MReg* on a large in-house A4C-CDV dataset comprising 1868 cases with three graded regurgitation labels. Compared to other weakly supervised video anomaly detection and supervised classification methods, *MReg* demonstrated superior performance in MR diagnosis. Our code is available at: <https://github.com/ckdstz/MReg>.

Keywords: Mitral regurgitation · Color Doppler Echocardiography · Feature Mining · Mixture-of-Experts

* Zhe Liu, Yuhao Huang and Lian Liu contribute equally to this work.
Corresponding email: gzhongshan1986@163.com and xinyang@szu.edu.cn

1 Introduction

Mitral regurgitation (MR) is a prevalent valvular heart disease, with precise severity assessment essential for treatment and prognosis. Color Doppler echocardiography is the most common tool in MR diagnosis due to its non-invasive nature, offering a clear visualization of blood flow to sonographers for disease assessment. However, MR diagnosis heavily relies on the sonographer’s experience, introducing subjectivity, increasing the risk of misdiagnosis and limiting the feasibility of large-scale screening. Hence, developing an automated system for MR diagnosis is crucial to improving diagnostic objectivity and accuracy.

Deep learning has been widely applied in the analysis of valvular heart disease, including structural segmentation, parametric measurements [4,6,7], and disease detection [18,20,5,11], e.g., mitral/aortic/tricuspid regurgitation. However, these methods were not specifically designed for MR diagnosis, which may lead to suboptimal accuracy. Recently, Sadeghpour et al. [15] developed a multiparametric machine-learning method to measure MR-related parameters for grading based on images. However, their strategies heavily relied on keyframe selection, ignoring crucial dynamic information for accurate assessment.

Two studies [12,17] used the naive video-based convolutional neural network model to classify MR. Although effective, they utilized conventional video classification models that were not specifically optimized for MR features. Thus, these methods may struggle with complex cases, such as multi-jet or eccentric regurgitation. Figure 1 shows that moderate and severe MR with eccentricity have similar regurgitant areas to mild MR, which may confuse the above methods and lead to inaccurate prediction. Moreover, data imbalances can lead to significant variations in diagnostic accuracy across categories, e.g., higher misdiagnosis rates for moderate MR [15]. Hence, an automated framework to address the above challenges is highly desirable for accurate and robust MR diagnosis.

In this paper, we develop a novel regression-based framework named *MReg* to predict normal (Grade 0), mild MR (Grade 1), and moderate-severe MR (Grade 2), using A4C-CDV. We combine moderate and severe MR, as both require clinical intervention, unlike mild MR requiring only periodic screening. Based on three feature mining methods, our highlights are: **(1)** To our knowledge, this is the first end-to-end video-based MR-specific qualitative diagnostic model formulated as a regression task to capture category relationships. **(2)** We introduce *feature selection* and *amplification* designs to mimic the clinical process, driven by multiple instance learning (MIL). **(3)** We propose an Mixture-of-Experts (MoE)-based *feature summary* module to achieve category-level feature decoupling for handling complex cases. **(4)** Extensive experiments show that our method significantly outperforms existing strong video-based methods.

2 Methodology

Figure 2 shows the overview of our proposed framework. Videos are first cropped to retain the region of interest (ROI), and sampled via MIL into three 16-

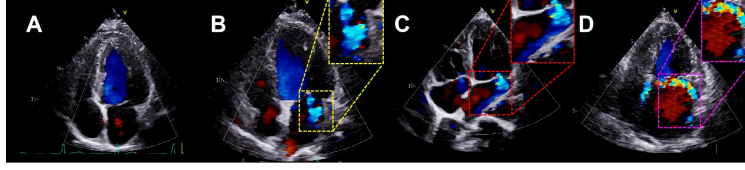


Fig. 1: Visualizations of normal (A) and various MR cases: (B) mild MR, (C)-(D) eccentric moderate and severe MR. Dashed boxes show the regurgitation jets.

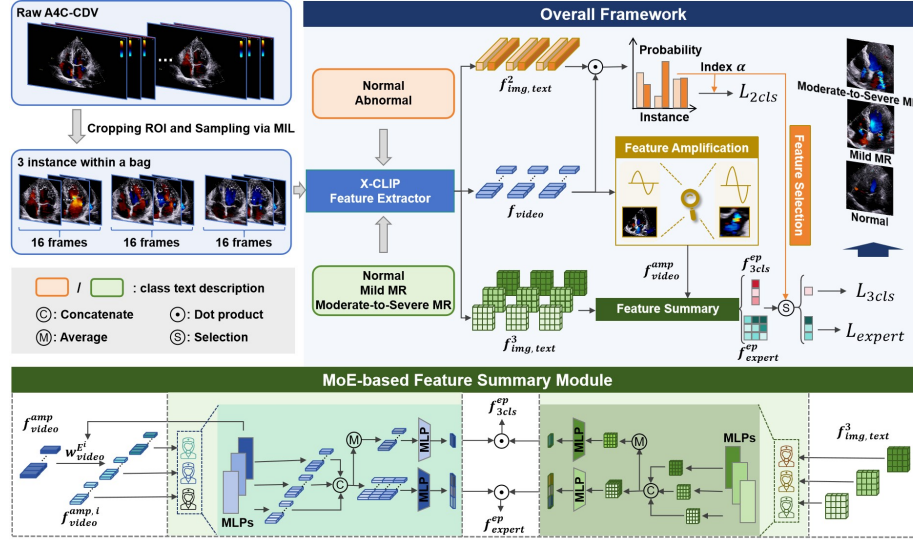


Fig. 2: Overview of our proposed method.

frame instances. The MIL-based operation is designed to facilitate the subsequent feature selection to use the most representative clip for diagnosis, avoiding interference from others. Then, X-CLIP [13] is used to extract video features $f_{video} \in R^{3 \times 16 \times 512}$, and two/three category text features: $f_{img, text}^2 \in R^{3 \times 2 \times 512}$ and $f_{img, text}^3 \in R^{3 \times 16 \times 196 \times 3 \times 512}$, respectively. Last, these features will be input into the following framework to perform efficient and accurate MR diagnosis.

Notably, previous methods often treat MR diagnosis as a classification task. However, this approach primarily suffers from the drawback that discrete category labels in classification scenarios may fail to capture the continuity and ordinal relationships between categories. Besides, the cross-entropy loss may further aggravate the learning of relative category relations, potentially breaking the learning stability. Hence, given the continuous spectrum of MR severity, we reformulate the MR diagnosis as a regression task and propose *MReg* to effectively model category relationships. Below are the method details, including three feature mining methods: feature selection, amplification, and summary.

2.1 Feature Selection and Amplification for Coarse MR Regression

Clinically, A4C-CDV-based MR diagnosis mainly contains two steps. First, sonographers determine the presence of MR and identify the cardiac cycle with the most severe regurgitation for severity assessment. Second, sonographers combine A4C-CDV local amplification video to observe the regurgitation area in more detail to achieve an accurate diagnosis. Inspired by this, we design a two-stage framework to match the clinical behavioral logic of MR analysis.

At stage I (i.e., **Step 1**), we follow the X-CLIP mechanism to determine the presence of MR and select the instance index $\alpha = \operatorname{argmax}(f_{2cls_out}[:, 1])$ with the highest MR probability for subsequent diagnosis. The focal loss [9] for binary classification $L_{2cls} = \operatorname{FocalLoss}(\operatorname{SoftMax}(S(f_{2cls_out}; \alpha)), \bar{Y}_2)$ supervises this process. f_{2cls_out} represents the model output, $S(\cdot)$ represents the selection of corresponding instance features according to α , and $\bar{Y}_2$ represents the 2-category label, i.e., normal or MR. At stage II (i.e., **Step 2**), f_{video} and $f_{img, text}^3$ will be used for MR diagnosis based on α . Here, we propose a feature amplification mechanism to amplify the features, which not only retains the complete cardiac motion information in A4C-CDV, but also amplifies the regurgitation feature information to facilitate MR diagnosis. The amplified feature f_{video}^{amp} is:

$$f_{video}^{amp} = \mathcal{N}(f_{video} + \beta \times \operatorname{Conv}(\|f_{video}\|_2)), \quad (1)$$

where $\mathcal{N}(\cdot)$ represents the feature normalization on the time and instance dimension to facilitate the subsequent calculation. β is the amplification coefficient. $\operatorname{Conv}(\cdot)$ represents a one-dimensional convolution to select the amplified features that are MR-relevant, and $\|\cdot\|_2$ indicates L_2 norm. The combination of amplified and original features not only retains the information extracted from the complete heart, but also adds effective local amplified information. It is noted that both features are extracted from the same video for convenient combination. Last, we use MSE loss [14] for model optimization:

$$L_{3cls} = \operatorname{MSELoss}(S(\mathcal{F}(f_{video}^{amp}, f_{img, text}^3); \alpha), \bar{Y}_3), \quad (2)$$

where $\mathcal{F}(\cdot)$ denotes dot product fusion after feature dimension transformation, and $\bar{Y}_3$ denote the 3-category label. We transform the model output into three categories based on thresholds $thre_1$ and $thre_2$. In this case, even if the regression model uses discrete labels, the continuity and order relationships of different categories can still be learned to enhance MR diagnosis.

2.2 Enhanced Regression via MoE-based Feature Summary Module

While the above feature selection and amplification can mimic the clinical workflow and yield good results, the complexity of the MR diagnostic mechanism still leads to unsatisfactory accuracy. Specifically, MR severity is influenced by various factors, such as regurgitation jet area, duration of regurgitation, and degree of eccentricity of regurgitation jet, and MR of varying severity typically contains different characteristics. Inspired by these clinical facts, we introduce

the MoE mechanism and built the feature summary module with three experts for $f_{img,text}^3$ and f_{video}^{amp} , respectively. Then, the proposed module aims to summarize MR features in different categories. Since $f_{img,text}^3$ has features of three categories and f_{video}^{amp} does not, we map f_{video}^{amp} into three directions to obtain $f_{video}^{amp,i}$, which is then summarized it by different experts:

$$w_{video}^{E^i} = SoftMax(\|W_{video}^i\|_F); f_{video}^{amp,i} = f_{video}^{amp} \odot w_{video}^{E^i}, \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm [2], and $W_{video}^i \in R^{512 \times 512}$ means the weight of each video expert. *SoftMax* function highlights relevant features by augmenting the weight of the most relevant expert and reducing the influence of irrelevant ones. The regression loss function of stage II is updated from (2) to $L_{3cls} = MSELoss(S(f_{3cls}^{ep}; \alpha), \bar{Y}_3)$, where f_{3cls}^{ep} denotes the model output after mixing three experts in the feature summary module. Additionally, we designed a focal loss based function to optimize the expert learning:

$$L_{expert} = FocalLoss(\argmax(S(f_{expert}^{ep}; \alpha)), \bar{Y}_3), \quad (4)$$

where f_{expert}^{ep} denotes the model output without mixing three experts. Therefore, with the smoothing L_{smooth} and sparsity losses $L_{sparsity}$, the total loss function can be written as: $L_{total} = L_{2cls} + L_{3cls} + L_{expert} + \lambda_1 L_{smooth} + \lambda_2 L_{sparsity}$, where λ_1 and λ_2 are two weight hyperparameters.

3 Experimental Results

Dataset and Implementations. The dataset comprised 1,868 A4C-CDV cases, categorized into Grade 0 (normal, 965), Grade 1 (mild MR, 677), and Grade 2 (moderate-severe MR, 226) by an experienced sonographer. Then, cases were randomly divided for training (450/296/103), validation (112/74/25), and testing (403/307/98). All frames in the videos were resized to 224×224 after ROI detection and MIL sampling. ROI is a rectangular area covering all cardiac chambers, obtained using YOLOv8 [16] pre-trained on our MR dataset with ROI labels annotated using Pair software [8]. The MIL sampling uses 3×16-frame non-overlapping clips per video. The data acquisition protocol ensured that each consecutive 16-frame clip encompassed more than 1 cardiac cycle.

All experiments were implemented in *PyTorch* (version 2.0.1) with two NVIDIA 3090 GPUs with 24G memory each. The training was optimized by AdamW in 100 epochs, with batch size=1 and learning rate=1e-5. We used X-CLIP pretrained on Kinetics-400 [3] for initialization. The amplification coefficient β was 2, and the thresholds $thre_1$ & $thre_2$ were 0.5 & 1.5, respectively. In the loss function, we set $\lambda_1=0.01$ and $\lambda_2=0.001$. We employed five evaluation metrics: Accuracy (%), Recall (%), Precision (%), Specificity (%), and F1-score (%). The model with the highest validation accuracy will be selected for evaluation.

Quantitative and Qualitative Analysis. We validated the effectiveness of the proposed *MReg* through comprehensive comparative experiments, ablation

Table 1: Method Comparison. The best results are shown in bold.

SW-VAD Methods	Accuracy	Recall	Precision	Specificity	F1-score	P-Value
AnomalyCLIP [21]	65.35	60.78	60.08	60.12	80.50	< .001
SVC Methods	Accuracy	Recall	Precision	Specificity	F1-score	P-Value
ViVit [1]	72.28	68.43	76.03	69.74	82.72	< .001
ActionCLIP [19]	82.67	82.36	81.28	81.47	90.86	< .001
Video-Swin [10]	86.76	85.89	82.94	84.22	93.18	< .001
<i>MReg</i>	89.36	85.93	86.83	86.36	94.28	-

Table 2: MR diagnostic performance of the model with different modules.

Model	Accuracy	Recall	Precision	Specificity	F1-score	P-Value
Baseline	82.67	75.57	81.31	77.73	90.09	< .001
Baseline + FS	87.38	83.45	84.42	83.90	93.25	< .001
Baseline + FS + Amp	88.74	84.87	85.60	85.23	93.98	< .001
Baseline* + FS + Amp	86.88	82.99	86.83	84.43	92.83	< .001
Baseline + FS + Amp + MoE	88.37	85.37	86.68	85.80	93.90	< .001
<i>MReg**</i>	81.06	74.87	81.46	77.13	88.59	< .001
<i>MReg</i>	89.36	85.93	86.83	86.36	94.28	-

* Corresponds to the *classification* model.

** Indicates the removal of FS in *MReg* and randomly selecting instances.

studies, and diverse visualizations. We also performed Chi-square test results between *MReg* and other methods for significance analysis.

Table 1 compares *MReg* with supervised video classification (SVC) and weakly supervised video anomaly detection (WS-VAD) methods. For SVC methods, we compare *MReg* with three widely used approaches, including ViVit [1], ActionCLIP [19] and VideoSwin [10]. WS-VAD-based methods are included since they share a similar mechanism with *MReg*, learning frame-level anomalies using video-level labels. However, since *MReg* not only detects anomalies (MR) but also categorizes them, we specifically consider AnomalyCLIP [21] with classification ability for comparison. The results demonstrate that *MReg* achieves state-of-the-art performance compared to all competing methods. Notably, WS-VAD methods perform significantly worse than SVC methods on all metrics. However, despite sharing a similar mechanism with WS-VAD, *MReg* significantly surpasses all SVC methods, further highlighting its superiority.

Table 2 reports the ablation studies for different proposed modules, while Figure 4 visualizes the corresponding results. We employ the X-CLIP-based single-stage regression model as the baseline. FS refers to the feature selection strategy, Amp denotes the feature amplification mechanism, MoE presents the feature summarization module, and *MReg* signifies the final diagnostic model incorporating the loss L_{expert} . Results validate the effectiveness of regression-based MR diagnosis compared to classification-based solution, particularly for MR samples (rows 3-4, Table 2, Recall: 82.99→84.87). Both FS and Amp significantly enhance diagnostic performance. The comparison between *MReg***

GT	Grade 0	93.05	6.95	0.00	Grade 0	92.56	7.44	0.00	Grade 0	94.79	5.21	0.00	Grade 0	94.29	5.71	0.00
	Grade 1	17.92	77.52	4.56	Grade 1	7.82	85.34	6.84	Grade 1	7.82	85.34	6.84	Grade 1	7.17	86.97	5.86
	Grade 2	0.00	43.88	56.12	Grade 2	0.00	27.55	72.45	Grade 2	0.00	25.51	74.49	Grade 2	0.00	23.47	76.53
		Baseline			Baseline+FS			Baseline+FS+Amp			MReg					

Fig. 3: Recall confusion matrix of the proposed method.

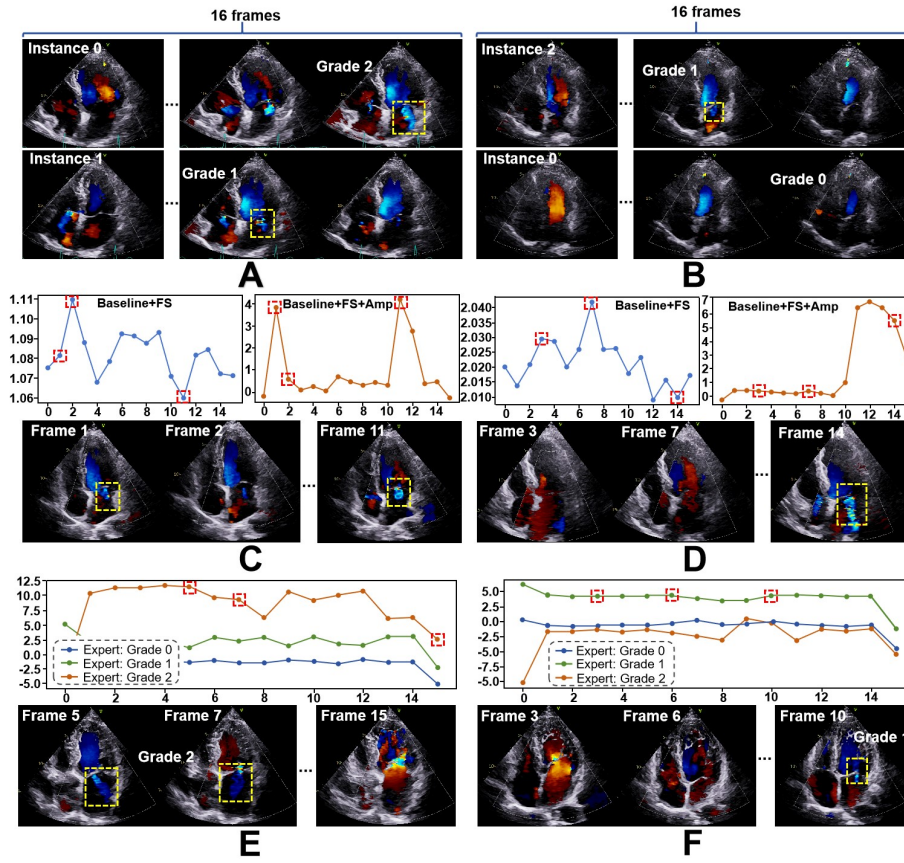


Fig. 4: **A** and **B** present instance selection for *MReg* (row 1) and *MReg*** (row 2) based diagnosis, with the index on the first image. **C-F** display per-frame regression results, including cases *w* and *w/o* Amp (C-D), and using feature summary (E-F). Instance labels are marked on the corresponding frames.

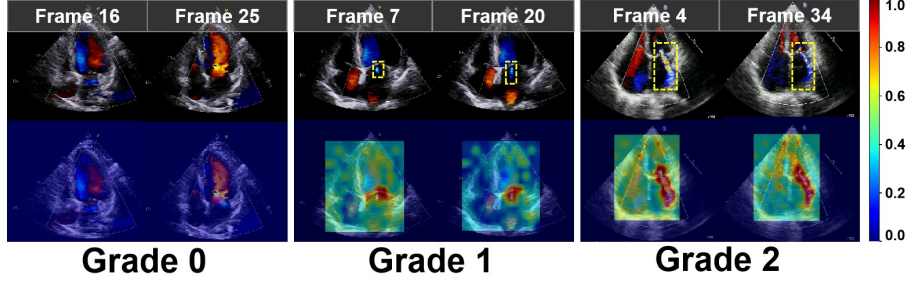


Fig. 5: The activation diagram of *MReg* in the three categories, with the original image at the top and the corresponding activation maps at the bottom.

and *MReg* further reveals the importance of accurate MR instance selection. Furthermore, as shown in Figure 3, our method has a stable improvement in both Grades 1-2 with less data, especially in Grade 2 (56.12 $\rightarrow$ 76.53, Baseline *vs.* *MReg*). While for Grade 0 with most data, *MReg* improves Baseline, and only degrades slightly compared to Baseline+FS+AMP.

It can be seen in Figure 4 (A)-(B) that both *MReg*^{**} cases resulted in incorrect judgments due to wrong instance selection. For example, in Figure 4 (A), we provide two instances of the same video. *MReg* accurately finds instance 0 with the most severe regurgitation and predicts the right category, i.e., Grade 2. However, *MReg*^{**} selects instance 1 lacking key regurgitation information, thus making the wrong decision (Grade 1). The results in Figures 4 (C) and (D) demonstrate that the regurgitation frame becomes more prominent after incorporating feature amplification (Amp), leading to more accurate predictions (as in Frames 1 and 11 in C). Without Amp, Baseline+FS assigns low scores to key abnormal frames, resulting in prediction errors. Furthermore, the gap between decision frames and others in the same instance increases after applying Amp (as largest gaps: ~ 0.05 *vs.* ~ 4 in C), making the decision frames dominant in the classification of the instance, reducing the influence of irrelevant frames.

Adding MoE improves the detection of positive samples, however, some metrics slightly decline (Accuracy/F1, Table 2). We attribute this to the absence of a supervision mechanism to guide the learning direction of the added expert models. This can be supported by the improved performance observed with L_{expert} . Figure 4 (E) and (F) display the regression results for per frame in two instances after applying the feature summary module. The results reveal that the regression value of the expert model for the corresponding category is higher, indicating that different experts effectively learn distinct categories, consistent with our design assumption. The activation maps in Figure 5 further validate that our *MReg* can accurately capture the vital regurgitation information. Furthermore, both sets of instances in Figure 4 (C)-(F) show that the regression values increase with the severity of MR. The increased gap between different MR categories after applying Amp facilitates more precise MR differentiation.

4 Conclusion

We propose a novel A4C-CDV-based MR diagnostic model, *MReg*, to categorize normal, mild, and moderate-severe MR. We first design feature selection and amplification strategies to mimic the diagnostic logic of sonographers. Then, we propose a MoE-based feature summary module that enables different experts to summarize feature information in different categories for more accurate diagnosis. Experiments on our in-house dataset validate the effectiveness of our approach. In the future, we will expand each MR subcategory to improve performance, and develop a multi-view diagnostic method.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (Nos. 82201851, 12326619, 62171290), Guangxi Province Science Program (No. 2024AB17023), Yunnan Major Science and Technology Special Project Program (No. 202402AA310052), Yunnan Key Research and Development Program (No. 202503AP140014), Suzhou Gusu Health Talent Program (Nos. GSWS2022071, GSWS2022072), Suzhou Key Laboratory of Cardiovascular Disease (No. SZS2024015) and Science and Technology Planning Project of Guangdong Province (No. 2023A0505020002).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., Schmid, C.: Vivit: A video vision transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6836–6846 (2021)
2. Böttcher, A., Wenzel, D.: The frobenius norm and the commutator. *Linear algebra and its applications* **429**(8-9), 1864–1885 (2008)
3. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
4. Chandra, V., Sarkar, P.G., Singh, V.: Mitral valve leaflet tracking in echocardiography using custom yolo3. *Procedia Computer Science* **171**, 820–828 (2020)
5. Cheng, L.H., Bosch, P.B., Hofman, R.F., Brakenhoff, T.B., Bruggemans, E.F., van der Geest, R.J., Holman, E.R.: Revealing unforeseen diagnostic image features with deep learning by detecting cardiovascular diseases from apical 4-chamber ultrasounds. *Journal of the American Heart Association* **11**(16), e024168 (2022)
6. Herz, C., Pace, D.F., Nam, H.H., Lasso, A., Dinh, P., Flynn, M., Cianciulli, A., Golland, P., Jolley, M.A.: Segmentation of tricuspid valve leaflets from transthoracic 3d echocardiograms of children with hypoplastic left heart syndrome using deep learning. *Frontiers in Cardiovascular Medicine* **8**, 735587 (2021)
7. Li, H., Wang, Y., Qu, M., Cao, P., Feng, C., Yang, J.: Echoefnet: multi-task deep learning network for automatic calculation of left ventricular ejection fraction in 2d echocardiography. *Computers in Biology and Medicine* **156**, 106705 (2023)
8. Liang, J., Yang, X., Huang, Y., Li, H., He, S., Hu, X., Chen, Z., Xue, W., Cheng, J., Ni, D.: Sketch guided and progressive growing gan for realistic and editable ultrasound image synthesis. *Medical image analysis* **79**, 102461 (2022)

9. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
10. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3202–3211 (2022)
11. Liu, Z., Zhu, X., Han, T., Huang, Y., Wang, J., Liu, L., Wang, F., Ni, D., Gou, Z., Yang, X.: Mitral regurgitation recognition based on unsupervised out-of-distribution detection with residual diffusion amplification. In: International Workshop on Machine Learning in Medical Imaging. pp. 52–62. Springer (2024)
12. Long, A., Haggerty, C.M., Finer, J., Hartzel, D., Jing, L., Keivani, A., Kelsey, C., Rocha, D., Ruhl, J., vanMaanen, D., et al.: Deep learning for echo analysis, tracking, and evaluation of mitral regurgitation (delineate-mr). *Circulation* (2024)
13. Ni, B., Peng, H., Chen, M., Zhang, S., Meng, G., Fu, J., Xiang, S., Ling, H.: Expanding language-image pretrained models for general video recognition. In: European conference on computer vision. pp. 1–18. Springer (2022)
14. Ren, J., Zhang, M., Yu, C., Liu, Z.: Balanced mse for imbalanced visual regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7926–7935 (2022)
15. Sadeghpour, A., Jiang, Z., Hummel, Y.M., Frost, M., Lam, C.S., Shah, S.J., Lund, L.H., Stone, G.W., Swaminathan, M., Weissman, N.J., et al.: An automated machine learning-based quantitative multiparametric approach for mitral regurgitation severity grading. *Cardiovascular Imaging* **18**(1), 1–12 (2025)
16. Terven J, Córdova-Esparza D M, R.G.J.A.: A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction* **5**(4), 1680–1716 (2023)
17. Vruthula, A., Duffy, G., Vukadinovic, M., Liang, D., Cheng, S., Ouyang, D.: High-throughput deep learning detection of mitral regurgitation. *Circulation* **150**(12), 923–933 (2024)
18. Wahlang, I., Maji, A.K., Saha, G., Chakrabarti, P., Jasinski, M., Leonowicz, Z., Jasinska, E.: Deep learning methods for classification of certain abnormalities in echocardiography. *Electronics* **10**(4), 495 (2021)
19. Wang, M., Xing, J., Liu, Y.: Actionclip: A new paradigm for video action recognition. *arXiv preprint arXiv:2109.08472* (2021)
20. Yang, F., Chen, X., Lin, X., Chen, X., Wang, W., Liu, B., Li, Y., Pu, H., Zhang, L., Huang, D., et al.: Automated analysis of doppler echocardiographic videos as a screening tool for valvular heart diseases. *Cardiovascular Imaging* **15**(4), 551–563 (2022)
21. Zanella, L., Liberatori, B., Menapace, W., Poiesi, F., Wang, Y., Ricci, E.: Delving into clip latent space for video anomaly recognition. *Computer Vision and Image Understanding* **249**, 104163 (2024)