



A fully convolutional neural network for the quantification of mitral regurgitation in echocardiography

Lu Zhong^{1#}, Qing Deng^{1#}, Yin Wang², Hongning Song¹, Jinling Chen¹, Qing Zhou¹, Jinsheng Xiao², Sheng Cao¹

¹Department of Ultrasound Imaging, Renmin Hospital of Wuhan University, Wuhan, China; ²Electronic Information School, Wuhan University, Wuhan, China

Contributions: (I) Conception and design: L Zhong, Q Deng, S Cao; (II) Administrative support: H Song, J Chen, Q Zhou, J Xiao, S Cao; (III) Provision of study materials or patients: H Song, J Chen, Q Zhou, S Cao; (IV) Collection and assembly of data: L Zhong, Q Deng, Y Wang; (V) Data analysis and interpretation: L Zhong, Q Deng, Y Wang, S Cao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Sheng Cao, MD. Department of Ultrasound Imaging, Renmin Hospital of Wuhan University, 238# Jiefang Road, Wuhan 430060, China. Email: caosheng209@126.com.

Background: Mitral regurgitation (MR) is the most common form of valvular heart disease (VHD), and the accurate assessment of MR severity is critical for clinical management. However, the quantitative assessment of MR is intricate and time-consuming, posing challenges for physicians in ensuring the precision of the results. Thus, our objective was to create an automated and reproducible artificial intelligence (AI) system. This study aimed to assist physicians in grading MR severity using color Doppler echocardiograms through the implementation of a fully convolutional neural network (FCN).

Methods: A retrospective cohort was established comprising 433 patients diagnosed with MR based on clinical criteria. Following screening, 269 patients met the inclusion criteria for the study. In total, 4,104 frames from apical 4-chamber view color Doppler flow images constituted the training and validation set, while 1,060 frames comprised the test set. Using the FCN, the MR flow convergence region was captured and segmented. The algorithm also estimated the parameter radius, which was employed to compute the effective regurgitant orifice area (EROA) and regurgitant volume (RV) based on the proximal isovelocity surface area. These measurements were subsequently graded following the 2017 American Society of Echocardiography (ASE) guidelines. The segmentation and grading performance of the model were assessed. Additionally, the diagnostic performance of the AI model was compared to that of ultrasound physicians with varying years of experience.

Results: In groups I, II, III, and IV, the rates of correctly identifying the radius were 0.56, 0.83, 0.86, and 0.89, while the grading accuracy was 0.95, 0.89, 0.88, and 0.91, respectively. Regarding patients with MR of different etiologies, the grading accuracy for the functional MR and degenerative MR groups was 0.82 and 0.90, respectively. Using Carpentier classification of MR as the criterion, the accuracy for groups I, II, and IIIb was 0.80, 0.90, and 0.83, respectively.

Conclusions: The model showed commendable performance, streamlining the clinical diagnostic process and enhancing the precision and stability of quantitative MR assessment.

Keywords: Mitral regurgitation (MR); fully convolutional neural network (FCN); color Doppler echocardiography; mitral regurgitation grading (MR grading)

Submitted Apr 10, 2024. Accepted for publication Sep 23, 2024. Published online Nov 11, 2024.

doi: 10.21037/qims-24-735

View this article at: <https://dx.doi.org/10.21037/qims-24-735>

Introduction

Mitral regurgitation (MR) is the most common valvular heart disease (VHD) (1). The National Heart, Lung, and Blood Institute (NHLBI) reported that the prevalence of MR in the overall population is approximately 1.7%, and increases with age, reaching up to 9.3% in individuals aged over 75 years (1). Goel *et al.* reported that the population prevalence of mild (grade I), moderate (grade II), moderate-to-severe (grade III), and severe (grade IV) MR was 19.2%, 1.6%, 0.3%, and 0.2%, respectively (2). Patients with mild MR can be asymptomatic for a long time and have a better prognosis than patients with more severe MR. Conversely, patients with severe MR may experience pulmonary hypertension, atrial fibrillation, heart failure, and even death (3). When determining treatment strategies, both the 2021 European Society of Cardiology and 2020 American College of Cardiology and American Heart Association guidelines for VHD management (4,5) emphasize the significance of the precise grading of MR. Therefore, the precise evaluation of MR severity is essential for diagnosis, treatment, and prognosis.

As a pivotal tool in the diagnosis and assessment of MR, transthoracic echocardiography (TTE) offers the advantages of being non-invasive, radiation free, cost effective, and easy to perform. In the quantitative assessment of valvular regurgitation via TTE, three principal methodologies are employed: pulsed wave Doppler quantification; volumetric quantification; and the proximal isovelocity surface area (PISA) method (6). Of these three methods, the PISA method is the most frequently used (6). With advancements in technology and society, the number of demands for echocardiogram has increased significantly, but the number of standardized trained ultrasound physicians is limited, and the deficiencies in the MR assessment process persist. These challenges include the process of dealing with multiple ultrasound sections, numerous parameters for analysis, difficulty in making comprehensive judgments involving qualitative, semi-quantitative, and quantitative parameters, as well as the poor reproducibility of crucial quantitative parameters like the measurement of the effective regurgitant orifice area (EROA) and regurgitant volume (RV) using the PISA method. These challenges hamper the ability of

clinicians to obtain accurate echocardiogram assessment results in a timely manner. Consequently, enhancing the accuracy and efficiency of MR grading is an urgent clinical issue. Any improvements in this area would be beneficial for patients diagnosed with MR, and could alleviate the workload of ultrasound physicians, and guide clinical diagnosis and treatment.

Recently, convolutional neural network (CNN) technology has experienced rapid development and has been extensively applied in the medical field, demonstrating high sensitivity and specificity (7-9). Predominant applications in echocardiography practice include image recognition and segmentation, image quality assessment, the measurement of left ventricular volume and function, and disease diagnosis (10-13). These applications contribute to enhancing diagnostic efficiency, accuracy, and consistency. Artificial intelligence (AI) also has numerous applications in the diagnosis of MR. Edwards *et al.* developed a machine-learning model that uses echocardiographic videos and images for view classification and MR detection, which achieved high accuracy (14). Similarly, Kwon *et al.* developed an AI algorithm that effectively detects MR through electrocardiograms (15). Further, advancements have been made in diagnostics based on heart sounds. For instance, the support vector machine system proposed by Maglogiannis *et al.* and the maximum likelihood binary splitting method proposed by Safara *et al.* have both achieved high accuracy in classifying cardiac valvular diseases (16,17). These developments indicate the potential and promising future of AI in the diagnosis of cardiac valvular diseases. Nevertheless, previous AI-assisted automated evaluations in echocardiography-based MR studies have primarily relied on image features, single parameters, or non-quantitative parameters. Consequently, further research needs to be conducted to explore the application of CNNs in the automatic measurement of quantitative assessment parameters in MR echocardiography.

Based on the above, this study formulated and evaluated a CNN framework for the automated analysis of color Doppler echocardiograms and the quantification of MR severity grading parameters. We present this article in accordance with the TRIPOD+AI reporting checklist (available at <https://qims.amegroups.com/article/>

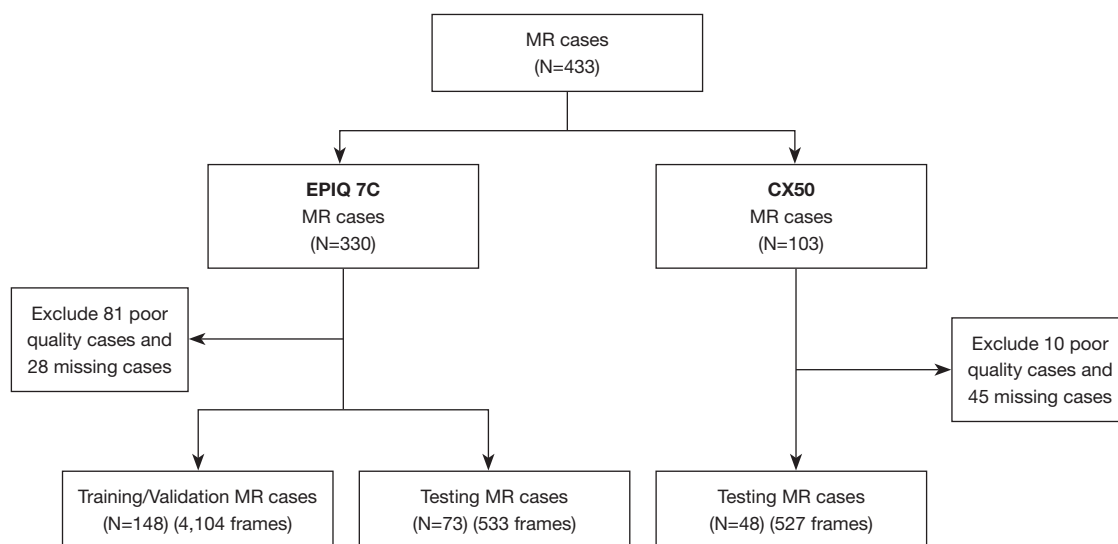


Figure 1 Flow chart for data selection, showing patient selection for training, validation, and testing. MR, mitral regurgitation.

view/10.21037/qims-24-735/rc).

Methods

Study population

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (No. WDRY2023-K186), and the requirement of individual consent for this retrospective analysis was waived. This study, which undertook an in-depth analysis of the model's application in a specific population, was conducted exclusively in secondary care settings. A total of 433 patients with MR from Renmin Hospital of Wuhan University from August 2021 to June 2024 were identified for inclusion in the study. After excluding cases with compromised image quality—16 with motion artifacts or acoustic shadows, 43 with inadequate frame rates, resolution, or inappropriate Nyquist limits, 12 featuring multiple jets of regurgitation, and 20 with indistinct hemispherical contours—along with 73 cases with incomplete images, the study ultimately included 269 patients. A total of 5,164 frames from apical 4-chamber color Doppler flow imaging of mitral valve (A4C-MV-CDFI) PISA flow convergence images were screened. Of these, 4,104 frames from 148 patients were used for training and validation purposes. Moreover, to assess the accuracy of “r” detection, 1,060 frames were selected from the remaining 121 cases in the test set. Additionally, one

frame was manually chosen from each case in the test set to evaluate grading accuracy (Figure 1).

Echocardiography

Eight experienced Asian ultrasound physicians (three males and five females, aged between 30 and 55 years), each with more than 10 years of experience and comprehensive specialized training, acquired all the echocardiograms in the study. All the echocardiographic images of the MR patients were obtained using EPIQ 7C (Philips Medical Systems, Best, The Netherlands) and portable CX50 (Philips Medical Systems, Best, The Netherlands) color Doppler ultrasound diagnostic machines. Each case involves the storage of four to six segments of dynamic images, with each segment containing three to five cardiac cycles. Color Doppler and continuous wave (CW) Doppler echocardiography were obtained in TTE apical 4-chamber (A4C) views. Images were stored and downloaded in the standard Digital Imaging and Communications in Medicine (DICOM) format. The echocardiographic images were reviewed for each case using Philips DICOM viewer software (version 3.0).

For this study, the EROA and RV were chosen as the quantitative assessment parameters. A crucial step involved measuring the regurgitant orifice radius (r) using the PISA method. The PISA method computation relies on the principles of hydrodynamics. The regurgitant blood flow, traversing a narrow regurgitant orifice, generates an

accelerated hemispherical blood flow region characterized by an escalating flow velocity and a diminishing surface area. A color flip transpires when the blood flow velocity surpasses the Nyquist limit. In accordance with the continuity equation, the counterflow rate through the counterflow port equals the counterflow rate through any isovelocity spherical surface. Consequently, the flow rate through the regurgitant orifice ($2\pi r^2 \times V_a$) can be computed to derive the EROA and RV. Simultaneously, manual measurements are taken of the maximum regurgitant velocity (V_{max}) and the velocity-time integral (VTI) at the valve orifice using CW Doppler spectroscopy. Subsequently, the EROA and RV were calculated using the equations $EROA = 2\pi r^2 \times V_a / V_{max}$ and $RV = EROA \times VTI$, respectively. The PISA methods are applicable across different sociodemographic groups, ensuring the comparability of the data and the fairness of the results.

Videos were selected for inclusion in the study if they at least met the following inclusion criteria: (I) the A4C color Doppler MR was fully included in the color sampling frame, and the three parts of the MR jet (i.e., the flow convergence, vena contracta, and regurgitant jet) were clear; and (II) the transducer frequency was appropriate, the color gain was optimized just below the clutter noise level, and the baseline of Nyquist limit was reasonably adjusted at 30–40 cm/s for PISA. These videos were reviewed frame by frame, and if a frame met the inclusion criteria, it was exported to a specific folder and numbered. Following the recommendations in the 2017 American Society of Echocardiography (ASE) guidelines (6), the severity-related indicators of MR, including the EROA and RV, were measured and categorized into the following four grades by the experienced ultrasound physicians: grade I, grade II, grade III, and grade IV.

Establishment and validation of the model

Image pre-processing

The LabelMe software (version 3.10, developed by MIT, Cambridge, MA, USA) was employed to annotate the region of interest, specifically the left ventricle lateral color region during systolic A4C-MV-CDFI in the MR images. Subsequently, the labeled files were input into an automated analysis network architecture to identify the region of blood flow convergence. The pre-processing measures included resizing the images to 512 pixels while maintaining the aspect ratio. Additionally, standardization, normalization, and random horizontal flipping techniques were employed.

These steps adjusted the data distribution and increased data diversity, thereby enhancing the performance and stability of the model.

Model training

In this study, the Deeplabv3+ fully convolutional neural network (FCN) model (*Figure 2*), a special type of CNN, was employed to identify the region of blood flow convergence at the MR mitral orifice. An overview of the proposed model is provided in *Figure 3*.

The K-means algorithm was applied to cluster the pixel points and color them according to the label value. This created a masked image. The pixel points were sorted according to their red, green, blue values, and the number of yellow pixel points in each row were counted, from top to bottom, and left to right. The highest point and the position with the fewest yellow pixels in the middle of the image was identified. This region refers to the features of the regurgitation convergence region that the model recognized. After multiple training sessions, our model accurately captured the image features of the regurgitation convergence region, and labeled the region to obtain the parameters necessary for quantitative MR evaluation. During the training process, each iteration used a batch size of four samples, with a total of 40,000 parameter updates. The parameter updates were performed using the stochastic gradient descent algorithm, with a momentum of 0.9 applied to accelerate convergence. To prevent overfitting, weight decay regularization was employed. The learning rate was adjusted using a polynomial decay strategy with an initial learning rate of 0.01, a final learning rate of 0, and a power of 0.9. For the classification task, the cross-entropy loss function was used to evaluate prediction errors, ensuring the accuracy and stability of the model.

Hardware specifications

This study was conducted using a computer equipped with the Linux Ubuntu 18.04.5 LTS operating system (developed by Canonical Ltd., London, UK) and an NVIDIA GeForce RTX 3090 (24GB) GPU (Graphics Processing Unit, developed by NVIDIA Corporation, Santa Clara, CA, USA). The construction, training, and image processing of the model primarily relied on paddlepaddle-gpu (version 2.4.2, developed by Baidu, Inc., Beijing, China) and paddleseg (version 2.8.0, developed by Baidu, Inc., Beijing, China). Data organization, analysis, and visualization were performed using pandas (version 1.3.5, developed by the Pandas Development Team), numpy (version 1.21.6), and

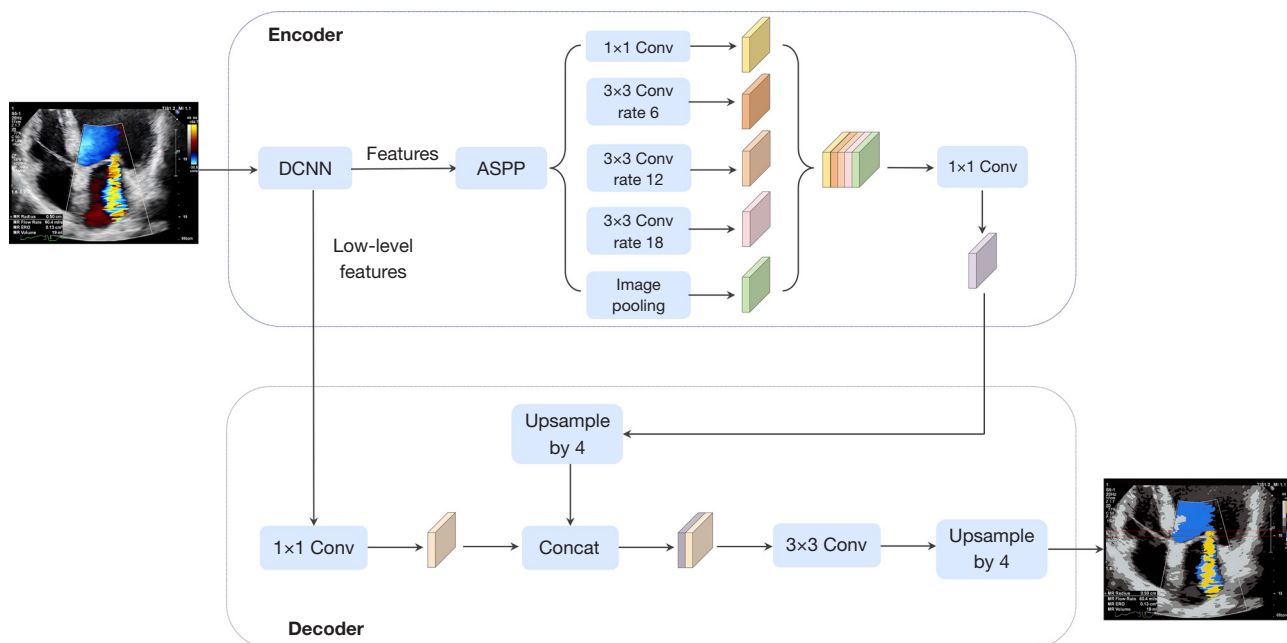


Figure 2 The network structure of the DeepLabv3+ neural network model. MR, mitral regurgitation; ERO, effective regurgitant orifice; DCNN, deep convolutional neural network; ASPP, atrous spatial pyramid pooling; Conv, convolution; A4C-MV-CDFI, apical 4-chamber color Doppler flow imaging of the mitral valve.

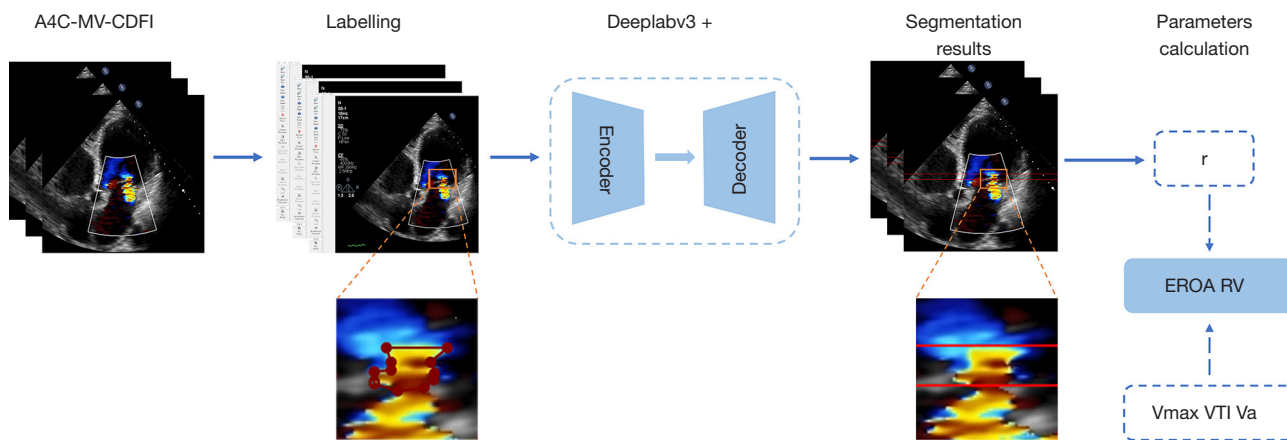


Figure 3 Detailed description of model architecture. The patients’ A4C-MV-CDFI data were used to quantify metric-associated lesion severity through pre-processing, contour extraction, and key point detection. The red dots and lines mark the areas of interest in the original image; the red horizontal lines represent the upper and lower bounds of the model output radius. A4C-MV-CDFI, apical 4-chamber color Doppler flow imaging of the mitral valve; EROA, effective regurgitant orifice area; RV, regurgitant volume; VTI, velocity-time integral.

opencv (version 4.5.5.64, developed by OpenCV.org).

Evaluation of the model

The data analysis was conducted using SPSS software

(version 26.0, developed by IBM Corporation, Armonk, NY, USA). The continuous variables are presented as the mean ± standard deviation (SD), median (interquartile range), count, or percentage. To evaluate the ability of the model to identify the regurgitant radius, manual evaluation was

Table 1 Baseline clinical and echocardiographic characteristics of the MR patients

Characteristic	Training/validation	Testing
Age (years)	66.3±10.1	64.8±12.3
Males (%)	54.2	65.8
Comorbidities (%)		
Coronary heart disease	55.6	58.0
Myocardial infarction	12.3	11.7
Hypertension	42.5	40.4
Diabetes	16.6	17.1
Atrial fibrillation	35.8	33.3
Echocardiography		
LAD (mm)	47 (43–51)	48 (44–52)
LVEDD (mm)	56 (48–60)	59 (53–62)
LVEF (%)	53 (38–56)	47 (40–56)
MR cases, n (frames)		
Grade I	31 (708)	19 (137)
Grade II	38 (1,281)	25 (230)
Grade III	36 (1,005)	30 (343)
Grade IV	43 (1,110)	47 (350)

Data are presented as the mean ± standard deviation, median (interquartile range), count, or percentage. MR, mitral regurgitation; LAD, left atrium anteroposterior diameter; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction.

used to determine the accuracy and non-detection rates for the test set. To avoid subjective bias, a double-blind method was employed in the assessment process. The evaluators remained completely blind to the personal information of the subjects during the outcome evaluation.

The overall performance of the AI model for MR grading was validated using accuracy, precision, recall, F1 score, a confusion matrix, and a Bland-Altman analysis. The Bland-Altman analysis was employed to compare the assessments of MR regurgitation severity by a highly experienced physician (Physician A), a less experienced physician (Physician B), and the FCN model.

Results

Baseline clinical and echocardiographic characteristics of the MR patients

In the training/validation dataset, the mean age of the MR patients was 66.3 years, 54.2% were male, and 55.6%, 12.3%, 42.5%, 16.6%, and 35.8% had coronary artery disease, myocardial infarction, hypertension, diabetes mellitus, and atrial fibrillation, respectively. In the test dataset, the mean age of the MR patients was 64.8 years, 65.8% were male, and 58.0%, 11.7%, 40.4%, 17.1%, and 33.3% had coronary artery disease, myocardial infarction, hypertension, diabetes mellitus and atrial fibrillation, respectively. The median of the left atrial anteroposterior diameter (LAD) and left ventricular end-diastolic diameter (LVEDD) were increased in the two datasets. The test dataset exhibited a lower left ventricular ejection fraction (LVEF) than the training/validation dataset. In the training/validation set, there were 31 cases of Grade I (708 frames), 38 of Grade II (1,281 frames), 36 of Grade III (1,005 frames), and 43 of Grade IV (1,110 frames). In the test set, there were 19 cases of Grade I (137 frames), 25 of Grade II (230 frames), 30 of Grade III (343 frames), and 47 of Grade IV (350 frames) (Table 1).

In the test dataset, 87 cases (797 frames) were classified as functional mitral regurgitation (FMR), while 34 cases (263 frames) were classified as degenerative mitral regurgitation (DMR). Using the Carpentier classification system as a standard, 58 cases (531 frames) were categorized as Type I, 34 cases (263 frames) as Type II, and 29 cases (266 frames) as Type IIIb.

Performance of model in flow convergence radius identification

Following model training, a visualization technique was used to visualize the learning process of the model (Figure 4). The green area in the overlaid image revealed the regions that the model prioritized to identify regurgitation convergence zones.

In the test dataset, the rates of correctly identified images for Grades I to IV were 0.56, 0.83, 0.86, and 0.89, respectively. Conversely, the rates of incorrectly identified

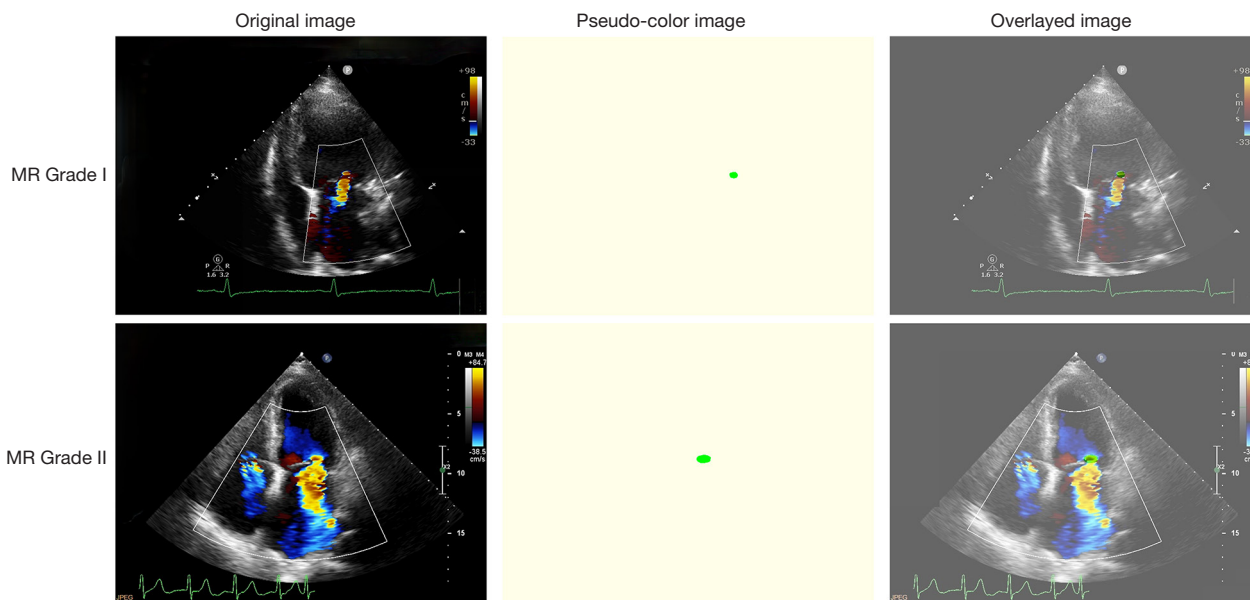


Figure 4 Visualization of the fully convolutional network model. MR, mitral regurgitation.

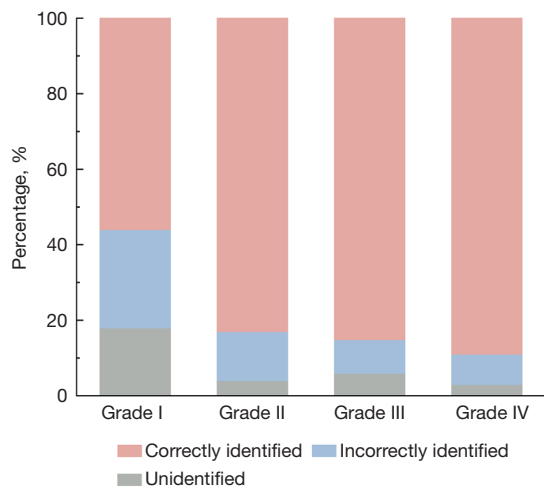


Figure 5 Performance of radius identification between groups on the testing data.

images for these grades were 0.26, 0.13, 0.09, and 0.08, respectively. Additionally, the rates of unidentified images were 0.18, 0.04, 0.06, and 0.03 for each corresponding grade (Figure 5). The accuracy of 0.56 for Grade I may be related to mild MR, for which the regurgitant area is limited, which poses challenges for radius recognition.

In the test dataset, the FMR group had a correctly identified rate of 0.86, while the DMR group achieved a rate of 0.92. The corresponding unidentified rates were 0.09

for FMR and 0.04 for DMR. The correctly identified rates for Type I, Type II, and Type IIIb were 0.84, 0.92, and 0.82, respectively, while the unidentified rates were 0.10, 0.04, and 0.12, respectively.

Ability of model to evaluate MR severity

In the test set, with reference to the grading results of the experienced physicians, the mean values for accuracy, precision, recall, and F1 score were approximately 0.91, 0.81, 0.83, and 0.81, respectively (Table 2). Figure 6 presents the confusion matrix for the grading results, illustrating that the majority of errors predominantly occurred between adjacent severity levels. Notably, in the internal test dataset, the predictive accuracy for Grades I and IV was higher than that for Grades II and III.

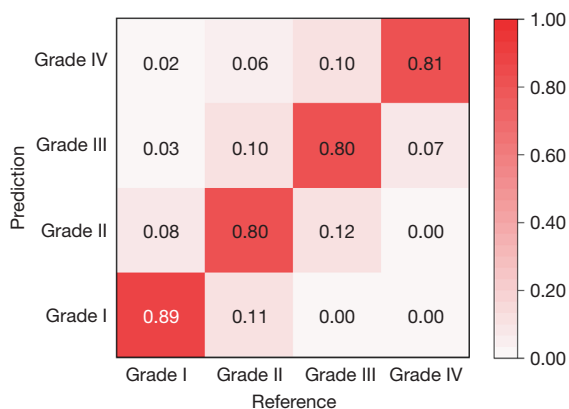
In the classification of MR etiology, the accuracy rates for the FMR and DMR groups were 0.82 and 0.90, respectively. In the Carpentier classification of MR, the accuracy rates for groups I, II, and IIIb were 0.80, 0.90, and 0.83, respectively.

Comparison of the performance of the physicians and AI model

Figure 7 shows the comparative results of the quantitative measurements for grading the severity of MR, specifically

Table 2 Predicted outcomes of mitral regurgitation severity

Grade	Accuracy	Precision	Recall	F1 score
Grade I	0.95	0.81	0.89	0.85
Grade II	0.89	0.71	0.80	0.75
Grade III	0.88	0.75	0.80	0.77
Grade IV	0.91	0.95	0.81	0.87
Mean	0.91	0.81	0.83	0.81

**Figure 6** The confusion matrix of the grading results of MR using the testing data. MR, mitral regurgitation.

the EROA and RV indices, between the FCN model and ultrasound physicians with varying years of experience. The mean difference obtained from the Bland-Altman analysis of the FCN model compared to the more experienced physicians (Group B) was slightly higher than that obtained from the analysis with the less experienced physicians (Group A). However, the range within ± 1.96 SD for the comparison between the FCN model and the more experienced physicians (Group B) was notably less, indicating a reduced dispersion of measurement discrepancies. Further, the data points predominantly clustered near the mean difference line, which showed that the predictive accuracy of the model approached that of the physicians with greater experience.

Discussion

In this study, we developed and tested an FCN framework for the automatic analysis of color Doppler echocardiograms. This framework was designed to quantify the severity grading parameters of MR by

leveraging the advanced capabilities of deep-learning algorithms to enhance diagnostic accuracy and efficiency. The FCN model was trained using a substantial dataset of echocardiographic images depicting cases of MR with varying degrees of severity. By automating the process of MR severity assessment, the framework aimed to reduce the inherent human errors and subjective variability present in traditional echocardiographic interpretations. The preliminary results showed that our FCN model achieved high accuracy and reliability in classifying the severity of MR. This approach not only simplifies the workflow of echocardiographic examinations but also serves as a valuable addition to the application of AI in cardiovascular imaging.

The PISA method is currently the most commonly used quantitative technique for assessing the degree of MR and plays a crucial role in determining treatment strategies and prognostic outcomes for patients (18-20). The challenge with the PISA method lies in accurately identifying the regurgitant radius on color Doppler images, which heavily depends on the skill of the ultrasound physician and is subject to significant interobserver variability. Additionally, the isovelocity surface area is calculated from the radius, resulting in errors that are magnified exponentially. According to the 2017 ASE guidelines (6), the values of the EROA and RV among Grades I to IV are very close, particularly for the EROA, where the difference between the grades is only 0.1 cm^2 , with values of <0.2 , $0.20\text{--}0.29$, $0.30\text{--}0.39$ and $>0.4 \text{ cm}^2$, respectively. This minimal variance can lead to measurement instability and directly affect clinical decision making, especially in determining whether a patient meets the criteria for transcatheter edge-to-edge repair or requires surgical valve replacement.

Previous studies have used AI to assess the severity of MR and have achieved some success. Yang *et al.* (12) used deep-learning algorithms to grade the severity of VHD based on automatic sectional classification and VHD diagnosis. These algorithms are capable of segmenting

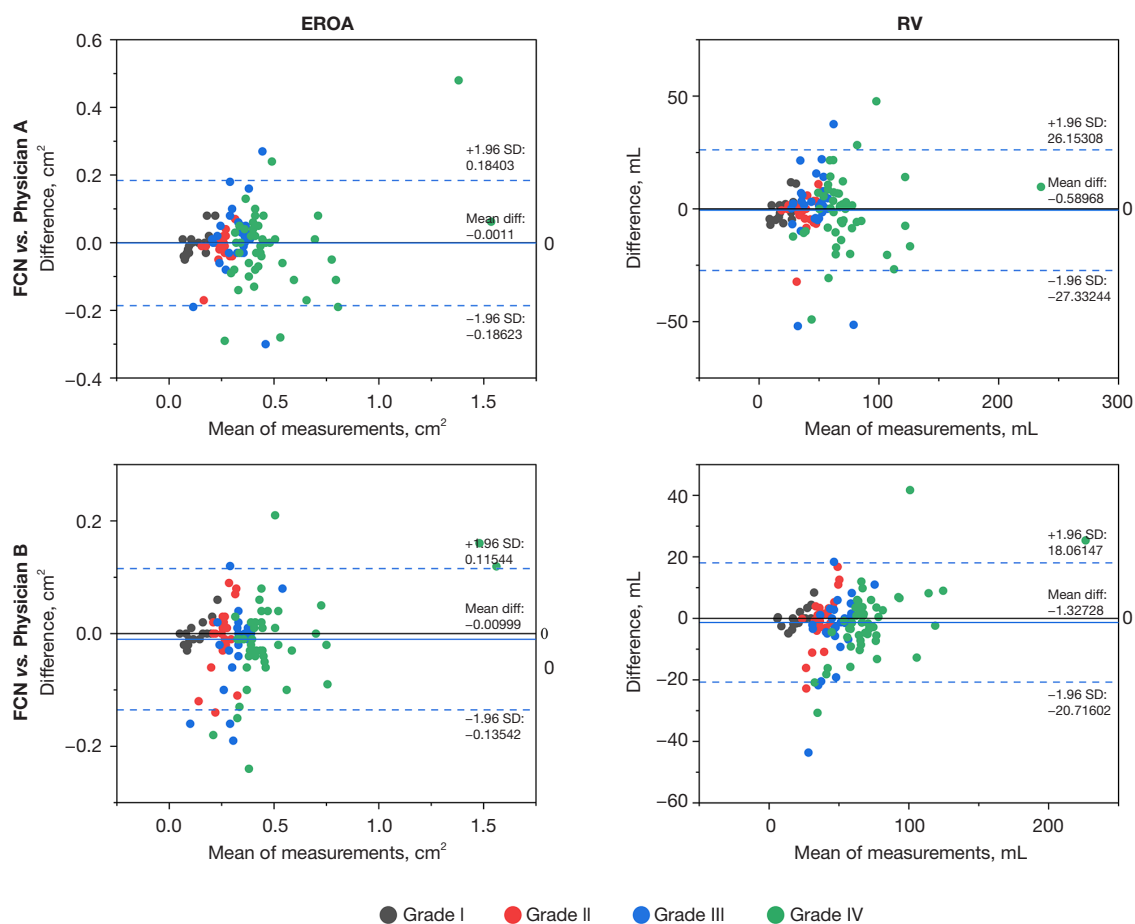


Figure 7 Comparisons of the quantitative metrics derived from the FCN algorithm, physician A, and physician B based on Bland-Altman analysis. The first and second rows compared the values computed after measurement by the FCN model with those of the less and more experienced physician, respectively. FCN, fully convolutional network; EROA, effective regurgitant orifice area; SD, standard deviation; RV, regurgitant volume.

key frames in color Doppler videos and quantifying semi-quantitative parameters related to MR severity, such as the MR jet area/left atrium (LA) area. Their capability in assessing MR severity was comparable to that of highly experienced ultrasound physicians. Additional research (21) introduced new image descriptors to better capture image features, selecting predictors such as the MR jet length/LA length, MR jet length, LA width, LA area, MR jet area and MR jet area/LA area, among which the MR jet length showed superior performance with an area under the curve of 0.953. Another study (22) employed CNN algorithms to assess MR severity, and achieved a classification accuracy of 90%, 87%, 81%, and 91% for Grades I to IV, respectively. Further, self-supervised learning algorithms based on two-dimensional images have also shown high sensitivity in

grading MR severity. These algorithms can automatically identify the frame with the largest regurgitant area and generate relevant parameters for assessing MR severity, thereby eliminating the variability of manual interpretation in clinical practice; however, they do not incorporate color Doppler image information (23). In more recent research, Tang *et al.* introduced a breakthrough approach called PISA-net, a fully automatic MR quantification method that includes processes such as cardiac cycle detection via electrocardiogram, Doppler spectrum segmentation, PISA radius segmentation based on M-mode echocardiography, and MR quantification. This method demonstrated a high Pearson correlation of up to 0.994 in measuring the RV and EROA, underscoring its potential and prospects for application in clinical MR diagnostics (24).

Compared to previous studies, this study introduced an innovative FCN model, which is the first to automatically assess the severity of MR using color Doppler images in conjunction with the PISA method. This model automatically and precisely measures the convergence radius, subsequently calculating the EROA and RV, and classifying the severity of MR in accordance with clinical guidelines. Additionally, the model includes two mutually validating parameters, which enhances the reliability of the assessment. The accuracy of this model in MR grading was comparable to that of experienced physicians. The algorithm not only improved the reproducibility and accuracy of diagnoses but also enhanced the comparability and interpretability of the results. This model can provide significant support for clinical decision making and patient management, and will contribute to the advancement of personalized medicine and precision therapy.

In the test results, the unidentified rate was higher for Grade I than for other grades. This is likely because the small regurgitant area in the color Doppler images of the Grade I MR patients made it difficult to identify the presence of MR. The accuracy rates for Grades II and III were slightly lower than those for Grades I and IV. This is likely because a small error of 1 mm near the threshold levels for Grade II and III MR could lead to changes in classification. Thus, MR classification should be combined with other indicators for comprehensive analysis. The accuracy of the FCN model-based PISA method in predicting the severity of patients with DMR and Carpentier Type II MR was higher than that of other groups for several reasons. First, for the same degree of MR, the PISA of FMR cases was significantly smaller than that of DMR cases, and the FCN model more easily recognized a larger convergence radius. Therefore, the smaller PISA radius in FMR cases might be more prone to measurement errors compared to the larger PISA radius in DMR cases. Additionally, the regurgitant orifices in DMR cases are mostly circular, while those in FMR cases are often non-circular or even elliptical, and circular regurgitant orifices and hemispherical PISA may be easier to identify. Further, the insufficient number of cases in the test set might have led to errors in the results; thus, further studies with increased sample sizes need to be conducted. The findings of this study only indicate that the FCN model can effectively measure the PISA in FMR and DMR cases.

The current FCN model still requires further validation and testing; however, there is reason to believe that it can serve as a supportive tool for ultrasound physicians

in interpreting MR images, particularly for those in economically underdeveloped regions or with limited experience. In addition to enhancing the diagnostic accuracy, it would also reduce the workload of physicians, especially in high-pressure clinical environments. Additionally, the deployment of this technology could serve as a critical component of quality control, ensuring adherence to the highest standards in the diagnostic process of MR. By analyzing a large volume of TTE data, the FCN system can identify potential grading anomalies, thereby prompting medical personnel to re-evaluate these specific cases. This contributes to the scientific rigor of MR grading assessments and improves the overall quality of medical services. At the same time, it must be recognized that deep-learning technologies cannot completely replace ultrasound physicians, especially in complex and critical diagnostic processes. Experienced ultrasound physicians rely not only on the images themselves but also integrate the overall clinical context of the patient, including medical history, symptoms, and results from other diagnostic tests. This comprehensive clinical judgment capability is currently difficult for deep-learning models to fully replicate.

Looking ahead, AI-based quantitative assessment technologies for MR are expected to expand to other types of VHD, particularly regurgitative disorders such as aortic valve regurgitation, tricuspid regurgitation, and pulmonary valve regurgitation. However, before these technologies can be widely adopted, they must be validated through diverse real-world clinical trials to verify their accuracy and efficiency to ensure that these methods will function stably and effectively in various clinical settings. In summary, deep-learning technologies should be viewed as powerful tools to enhance, but not replace, the diagnostic capabilities of physicians. These technologies should be used under the supervision of experienced medical professionals to improve healthcare efficiency and quality, while also ensuring that the highest commitment to patient safety and welfare is maintained. This collaborative model not only maximizes the potential of deep-learning technologies but also ensures that human doctors' expertise and care remain at the core of complex medical decision making.

This study had several limitations. First, our model excluded cases with poor quality images, which could lead to reduced performance when applied in real-world scenarios. Second, as mentioned above, the guidelines indicate that the assessment of MR severity should integrate multiple parameters. Our model incorporated two quantitative parameters; however, the inclusion of

more quantitative parameters could enhance the accuracy of grading, especially when two parameters do not align in their MR severity classification. Further, this was an initial study of a MR quantification model based on echocardiography; however, in the future, we plan to introduce automatic recognition of the MR regurgitation spectrum and the automatic selection of frames to fully automate the assessment process. Additional sensitivity analyses should be conducted to evaluate the potential impact of changes in the geometric assumptions about the regurgitant convergence area on model outputs, which we consider an important part of our future research plans. In this study, some data points appeared outside the ± 1.96 SD consistency limits. The presence of outliers suggests that, theoretically, the model may have demonstrated excellent predictive accuracy, but its performance in practice might be limited by the representativeness and diversity of the training data. Therefore, this underscores the importance of the continuous monitoring and evaluation of AI model performance, especially before models are widely deployed in clinical practice.

Conclusions

In this study, an FCN model was developed for the quantitative assessment of MR severity. The model demonstrated good performance, proving that the automated assessment of MR severity is feasible and effective. This model could assist clinicians in decision-making and enhance the accuracy of ultrasound physicians' assessments.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD+AI reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-735/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-735/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (No. WDRY2023-K186), and the requirement of individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet* 2006;368:1005-11.
2. Goel SS, Bajaj N, Aggarwal B, Gupta S, Poddar KL, Ige M, Bdair H, Anabtawi A, Rahim S, Whitlow PL, Tuzcu EM, Griffin BP, Stewart WJ, Gillinov M, Blackstone EH, Smedira NG, Oliveira GH, Barzilai B, Menon V, Kapadia SR. Prevalence and outcomes of unoperated patients with severe symptomatic mitral regurgitation and heart failure: comprehensive analysis to determine the potential role of MitraClip for this unmet need. *J Am Coll Cardiol* 2014;63:185-6.
3. Enriquez-Sarano M, Akins CW, Vahanian A. Mitral regurgitation. *Lancet* 2009;373:1382-94.
4. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, et al. 2021 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J* 2022;43:561-632. Erratum in: *Eur Heart J* 2022;43:2022.
5. Writing Committee Members; Otto CM, Nishimura RA, Bonow RO, Carabello BA, Erwin JP 3rd, Gentile F, Jneid H, Krieger EV, Mack M, McLeod C, O'Gara PT, Rigolin VH, Sundt TM 3rd, Thompson A, Toly C. 2020 ACC/AHA Guideline for the Management of Patients With Valvular Heart Disease: Executive Summary:

- A Report of the American College of Cardiology/ American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol* 2021;77:450-500. Erratum in: *J Am Coll Cardiol* 2021;77:1276.
6. Zoghbi WA, Adams D, Bonow RO, Enriquez-Sarano M, Foster E, Grayburn PA, Hahn RT, Han Y, Hung J, Lang RM, Little SH, Shah DJ, Shernan S, Thavendiranathan P, Thomas JD, Weissman NJ. Recommendations for Noninvasive Evaluation of Native Valvular Regurgitation: A Report from the American Society of Echocardiography Developed in Collaboration with the Society for Cardiovascular Magnetic Resonance. *J Am Soc Echocardiogr* 2017;30:303-71.
 7. Bong JH, Kim TH, Jeong S. Deep learning model for the diagnosis of breast cancers smaller than 1 cm with ultrasonography: integration of ultrasonography and clinical factors. *Quant Imaging Med Surg* 2023;13:2486-95.
 8. Chang X, Wang J, Zhang G, Yang M, Xi Y, Xi C, Chen G, Nie X, Meng B, Quan X. Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network. *Cell Rep Med* 2023;4:100914.
 9. Hou Y, Jia S, Lun X, Hao Z, Shi Y, Li Y, Zeng R, Lv J. GCNs-Net: A Graph Convolutional Neural Network Approach for Decoding Time-Resolved EEG Motor Imagery Signals. *IEEE Trans Neural Netw Learn Syst* 2024;35:7312-23.
 10. Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-Time Standard View Classification in Transthoracic Echocardiography Using Convolutional Neural Networks. *Ultrasound Med Biol* 2019;45:374-84.
 11. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;580:252-6.
 12. Yang F, Chen X, Lin X, Chen X, Wang W, Liu B, et al. Automated Analysis of Doppler Echocardiographic Videos as a Screening Tool for Valvular Heart Diseases. *JACC Cardiovasc Imaging* 2022;15:551-63.
 13. Sveric KM, Ulbrich S, Dindane Z, Winkler A, Botan R, Mierke J, Trausch A, Heidrich F, Linke A. Improved assessment of left ventricular ejection fraction using artificial intelligence in echocardiography: A comparative analysis with cardiac magnetic resonance imaging. *Int J Cardiol* 2024;394:131383.
 14. Edwards LA, Feng F, Iqbal M, Fu Y, Sanyahumbi A, Hao S, McElhinney DB, Ling XB, Sable C, Luo J. Machine Learning for Pediatric Echocardiographic Mitral Regurgitation Detection. *J Am Soc Echocardiogr* 2023;36:96-104.e4.
 15. Kwon JM, Kim KH, Akkus Z, Jeon KH, Park J, Oh BH. Artificial intelligence for detecting mitral regurgitation using electrocardiography. *J Electrocardiol* 2020;59:151-7.
 16. Maglogiannis I, Loukis E, Zafiroopoulos E, Stasis A. Support Vectors Machine-based identification of heart valve diseases using heart sounds. *Comput Methods Programs Biomed* 2009;95:47-61.
 17. Safara F, Doraisamy S, Azman A, Jantan A, Abdullah Ramaiah AR. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Comput Biol Med* 2013;43:1407-14.
 18. Enriquez-Sarano M, Miller FA Jr, Hayes SN, Bailey KR, Tajik AJ, Seward JB. Effective mitral regurgitant orifice area: clinical use and pitfalls of the proximal isovelocity surface area method. *J Am Coll Cardiol* 1995;25:703-9.
 19. Recusani F, Bargiggia GS, Yoganathan AP, Raisaro A, Valdes-Cruz LM, Sung HW, Bertucci C, Gallati M, Moises VA, Simpson IA, et al. A new method for quantification of regurgitant flow rate using color Doppler flow imaging of the flow convergence region proximal to a discrete orifice. An in vitro study. *Circulation* 1991;83:594-604.
 20. Enriquez-Sarano M, Avierinos JF, Messika-Zeitoun D, Detaint D, Capps M, Nkomo V, Scott C, Schaff HV, Tajik AJ. Quantitative determinants of the outcome of asymptomatic mitral regurgitation. *N Engl J Med* 2005;352:875-83.
 21. Yang F, Zhu J, Wang J, Zhang L, Wang W, Chen X, Lin X, Wang Q, Burkhoff D, Zhou SK, He K. Self-supervised learning assisted diagnosis for mitral regurgitation severity classification based on color Doppler echocardiography. *Ann Transl Med* 2022;10:3.
 22. Zhang Q, Liu Y, Mi J, Wang X, Liu X, Zhao F, Xie C, Cui P, Zhang Q, Zhu X. Automatic Assessment of Mitral Regurgitation Severity Using the Mask R-CNN Algorithm with Color Doppler Echocardiography Images. *Comput Math Methods Med* 2021;2021:2602688.
 23. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. *Comput Biol Med* 2016;73:47-55.
 24. Tang K, Ge Z, Ling R, Cheng J, Xue W, Pan C, Shu X,

Ni D. Mitral regurgitation quantification from multi-channel ultrasound images via deep learning. *Medical Image Computing and Computer Assisted Intervention*

– MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, Proceedings, Part VI, 2023:223-32.

Cite this article as: Zhong L, Deng Q, Wang Y, Song H, Chen J, Zhou Q, Xiao J, Cao S. A fully convolutional neural network for the quantification of mitral regurgitation in echocardiography. *Quant Imaging Med Surg* 2024;14(12):8707-8719. doi: 10.21037/qims-24-735