



OPEN A deep learning-based method for assessing tricuspid regurgitation using continuous wave Doppler spectra

Shenghua Xie^{1,2,11}, Han Liu^{3,11}, Li Su⁴, Jie Shen⁵, Junwang Miao⁶, Duo Huang⁷, Mi Zhou⁸, Huiruo Liu⁹, Yan Li¹⁰, Lixue Yin^{1,2}✉, Qinglan Shu^{1,2}✉ & Yi Wang^{1,2}✉

Transthoracic echocardiography (TTE) is widely recognized as one of the principal modalities for diagnosing tricuspid regurgitation (TR). The diagnostic procedures associated with conventional methods are intricate and labor-intensive, with human errors leading to measurement variability, with outcomes critically dependent on the operators' diagnostic expertise. In this study, we present an innovative assessment methodology for evaluating TR severity utilizing an end-to-end deep learning system. This deep learning system comprises a segmentation model of single cardiac cycle TR continuous wave (CW) Doppler spectra and a classification model of the spectra, trained on the TR CW Doppler spectra from a cohort of 11,654 patients. The efficacy of this intelligent assessment methodology was validated on 1500 internal cases and 573 external cases. The receiver operating characteristic (ROC) curves of the internal validation results indicate that the deep learning system achieved the areas under curve (AUCs) of 0.88, 0.84, and 0.89 for mild, moderate, and severe TR, respectively. The ROC curves of the external validation results demonstrate that the system attained the AUCs of 0.86, 0.79, and 0.87 for mild, moderate, and severe TR, respectively. Our study results confirm the feasibility and efficacy of this novel intelligent assessment method for TR severity.

Heart valvular disease is a prevalent cardiovascular condition, with an estimated 18–19% prevalence of valve regurgitation in middle-aged adults¹. This prevalence is expected to rise further due to the overall increase in the aging population². Currently, there are approximately 209 million patients worldwide suffering from valvular heart disease, with around 25 million cases in China alone³. TR is the most common type among various valvular diseases and is often associated with non-valvular heart disease. Accurate assessment of the severity of TR is an important step in determining the severity of the condition, evaluating treatment efficacy, and predicting prognosis. It plays an essential role in the diagnosis and treatment of cardiovascular diseases. The severity of regurgitation can be categorized into several grades: mild, moderate, severe, and more. Currently, TTE is the most commonly used noninvasive tool for assessing the severity of valve regurgitation^{4–7}. According to echocardiographic diagnostic guidelines and recommendations^{8–13}, the comprehensive evaluation of valve regurgitation severity is primarily based on quantitative indicators^{14–20}, such as the effective regurgitant orifice area (EROA), vena contracta width, the regurgitant volume (RVol), along with qualitative indicators such as

¹Ultrasound in Cardiac Electrophysiology and Biomechanics Key Laboratory of Sichuan Province, Sichuan Clinical Research Center for Cardiovascular Disease, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 610072, China. ²Department of Cardiovascular Ultrasound and Noninvasive Cardiology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 610072, China. ³School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu 610500, China. ⁴Department of Cardiovascular Medicine, Chengdu Fifth People's Hospital (The Second Clinical Medical College, Affiliated Fifth People's Hospital of Chengdu University of Traditional Chinese Medicine), Chengdu 611137, China. ⁵Department of Ultrasound Medicine, People's Hospital of Guangxi Zhuang Autonomous Region, Nanning 530021, China. ⁶Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Taiyuan 030032, China. ⁷Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, China. ⁸Department of Ultrasound in Medicine, Chengdu Wenjiang District People's Hospital, Chengdu 611130, China. ⁹Department of Ultrasound, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450000, China. ¹⁰Nanchong Central Hospital (Beijing Anzhen Hospital Nanchong Hospital), the Second Clinical College of North Sichuan Medical College, Nanchong 637003, China. ¹¹These authors contributed equally: Shenghua Xie and Han Liu. ✉email: yinlixue_cardiac@163.com; qinglanshu@163.com; wangyihdl@126.com

the CW Doppler spectrum. The assessment process requires multiple manual measurements and subjective qualitative evaluations. Therefore, traditional grading of valve regurgitation using TTE is not only time-consuming but also highly susceptible to inter- and intra-observer variability. The development of artificial intelligence has led to the emergence of intelligent evaluation methods for valve regurgitation based on deep learning networks^{14,21–26}, which show promise in addressing the aforementioned issues. These methods aim to achieve objective and precise assessment of valve regurgitation, enhance diagnostic efficiency, and meet the growing clinical demands.

In the field of valvular regurgitation assessment using deep learning methods, recent studies have demonstrated the potential of artificial intelligence technologies to improve diagnostic efficiency and accuracy. For instance, Zhang et al.²⁵ proposed the use of the Mask R-CNN algorithm to segment the mitral regurgitation areas in color Doppler echocardiography images. They utilized the segmented images to assess the severity of regurgitation based on information extracted from these images. Huang et al.²⁶ developed a framework based on VABC-UNet for automatic segmentation and classification of mitral and TR using deep learning methods. This approach not only enhanced image segmentation performance but also enabled automatic classification and grading of regurgitation as mild, moderate, moderately severe, or severe based on the characteristics of the regurgitation jet and the atrium. In a recent study, Wifstad et al.¹⁴ utilized three-dimensional Doppler ultrasound images and deep learning to quantify cardiac valvular regurgitation. The research enhanced the accuracy of estimating the orifice size and flow of regurgitation through deep learning networks, significantly improving the precision of area and flow measurements compared to traditional methods. However, existing intelligent assessment methods for valvular regurgitation typically require the selection of optimal images at key moments of regurgitation and involve complex procedural implementations. Therefore, considering the operability in clinical applications, we propose a deep learning-based approach to automatically classify the severity of TR using CW Doppler spectra. This study explores a simple and practical method for the intelligent assessment of the severity of TR during the echocardiography diagnose.

Results

A deep learning system has been meticulously developed for the intelligent assessment of TR severity using the raw TR CW Doppler spectra. An overview of the system's development, workflow, and test results is comprehensively illustrated in Fig. 1, providing clear insights into the system's development, workflow, efficacy.

Training results of the models for segmenting single cardiac cycle TR spectra

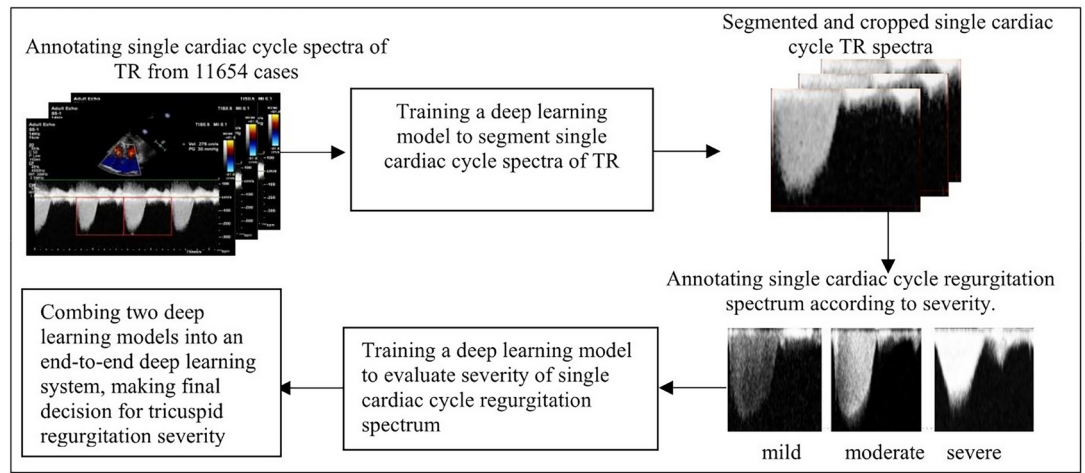
In order to ascertain the most effective deep learning model for the automated segmentation of single complete cardiac cycle regurgitation spectra, we meticulously annotated 15,345 images of these spectra. These images were randomly allocated into a training set and a test set in a 9:1 ratio. The segmentation models were trained utilizing the training set, and their performance was subsequently evaluated using the test set. Comparative experiments were executed using a range of prevalent object detection networks, including YOLOv5^{27,28} (YOLOv5l, YOLOv5s, YOLOv5m, YOLOv5x), YOLOv3, YOLOv6, Faster R-CNN, and SSD. Following 300 training epochs, we computed four essential performance evaluation metrics: Macro Average Precision (MAP), Macro Average Sensitivity (MAS), Macro Average F1 Score (MAF1), and mAP@0.5, as delineated in Table 1. The results revealed that the YOLOv5l network exhibited superior performance in terms of MAP, MAR, MAF1, and mAP@0.5, surpassing the other networks. Consequently, the YOLOv5l model demonstrated the most robust overall performance in this training, and we ultimately chose it as the model for segmenting single cardiac cycle regurgitation spectra.

Training results of the model for classifying single cardiac cycle TR spectra

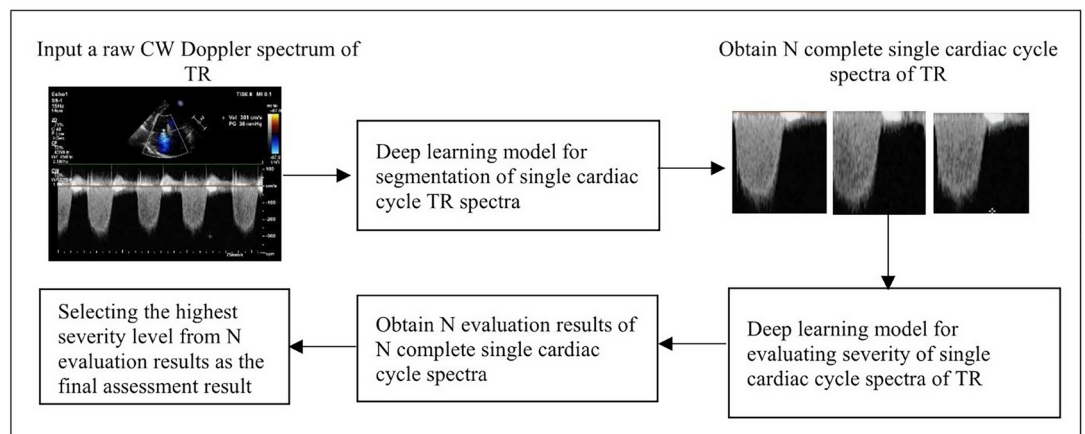
To determine the optimal deep learning model for classifying single cardiac cycle TR spectra by severity, we conducted a series of comparative experiments using multiple high-performance deep learning models. The evaluated deep learning models included ConvNeXt²⁹, RepVGG³⁰, Swin Transformer³¹, EfficientNetV2³², and ResNet50³³. We utilized a labeled dataset comprising 23,749 frames of single cardiac cycle regurgitation spectra, categorized into 7942 frames as mild, 8129 frames as moderate, and 7678 frames as severe. The dataset was divided into training, validation, and test sets in an 8:1:1 ratio. The classification models were trained using these aforementioned deep learning networks. After 300 training epochs, the performance of these models was assessed on test set using four evaluation metrics: Accuracy, MAS, MAP, and MAF1. The test results are detailed in Table 2. As illustrated in Table 2, the ConvNeXt model surpasses all other models across all evaluation metrics, demonstrating its superior overall performance. The detailed evaluation metrics for the ConvNeXt model on the test set are presented in Table 3. The confusion matrix for the ConvNeXt model on the test set is depicted in Fig. 2. The confusion matrix visually demonstrates the sensitivity of the classification model for each category, indicating that the primary prediction errors occur between adjacent severity levels. Figure 3 illustrates ROC curves of the ConvNeXt model on the test data. The ROC curves reveal that the AUCs of the three classifications exceeds 0.9, demonstrating the model's excellent performance in classifying single cardiac cycle TR spectra by severity. These findings confirm that the ConvNeXt model can provide high-quality assessments of regurgitation severity based on TR spectra. Consequently, the ConvNeXt model was ultimately selected as the model for classifying single cardiac cycle regurgitation spectra.

Test results of the deep learning system

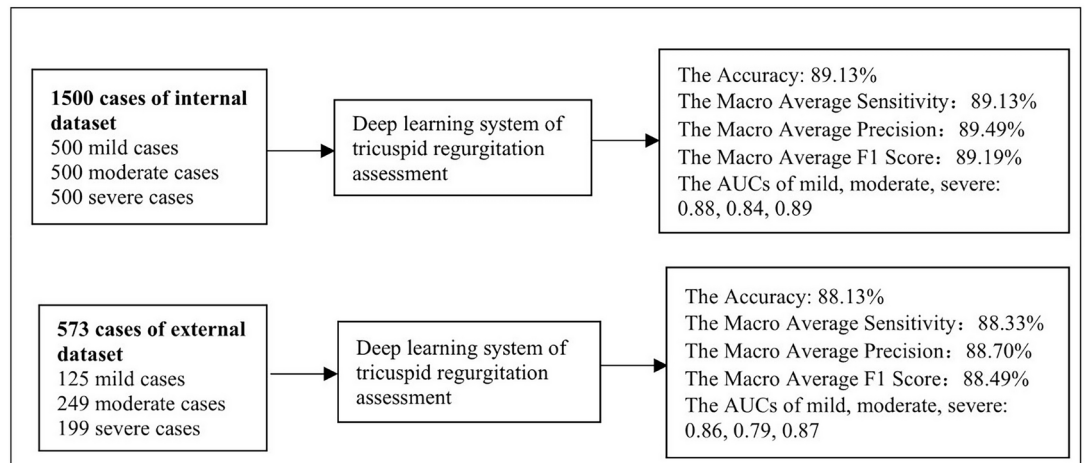
We have developed a deep learning system to predict TR severity by integrating the aforementioned single cardiac cycle spectrum segmentation model and TR severity classification model. The operational workflow of the deep learning system is shown in Fig. 1b. This system enables the input of a raw TR CW Doppler spectrum to generate an assessment result of TR severity. Therefore, this system is an end-to-end deep learning system



a The development process of the deep learning system



b The workflow of the deep learning system



c Test results

Fig. 1. Overview of the development of the deep-learning system, its workflow, and the test results. Subfigure (a) depicts the development process of the deep-learning system, which involves two steps: training a deep learning network to segment the single cardiac cycle regurgitation spectrum, and training another deep learning network to evaluate the severity of the regurgitation spectrum. Subfigure (b) outlines the workflow of the deep-learning system, indicating it as an end-to-end system. Subfigure (c) presents the results of the validation study, incorporating both internal dataset from our hospital and external dataset from seven other hospitals.

Model	MAP (%)	MAS (%)	mAP@0.5	MAF1 (%)
YOLOv5l	93.79	93.85	0.9625	93.82
YOLOv5s	92.91	93.11	0.9594	93.01
YOLOv5m	93.33	93.12	0.9621	93.22
YOLOv5x	92.41	93.19	0.9577	92.80
YOLOv3	92.53	92.98	0.9523	92.75
YOLOv6	92.69	93.01	0.9503	92.85
Faster R-CNN	92.26	94.88	0.9589	93.55
SSD	92.35	92.66	0.9491	92.50

Table 1. The performance comparison of YOLOv5 model with other main models.

Model	Accuracy (%)	MAS (%)	MAP (%)	MAF1 (%)
ConvNeXt	91.42	91.40	91.77	91.58
RepVGG	89.85	89.11	89.06	89.08
Swin Transformer	90.26	90.88	90.89	90.88
EfficientNetV2	90.35	90.66	90.91	90.78
ResNet50	87.53	87.98	87.23	87.60

Table 2. The performance comparison of the ConvNeXt model with other main models.

TR severity	Sensitivity (%)	Precision (%)	F1 score (%)	AUC
mild	90.93	94.88	92.86	0.98
moderate	92.47	85.50	88.85	0.91
severe	90.81	94.92	92.82	0.98
Macro average	91.40	91.77	91.58	0.96

Table 3. The evaluation metrics of the ConvNeXt model on the test date of a single cardiac cycle regurgitation spectra.

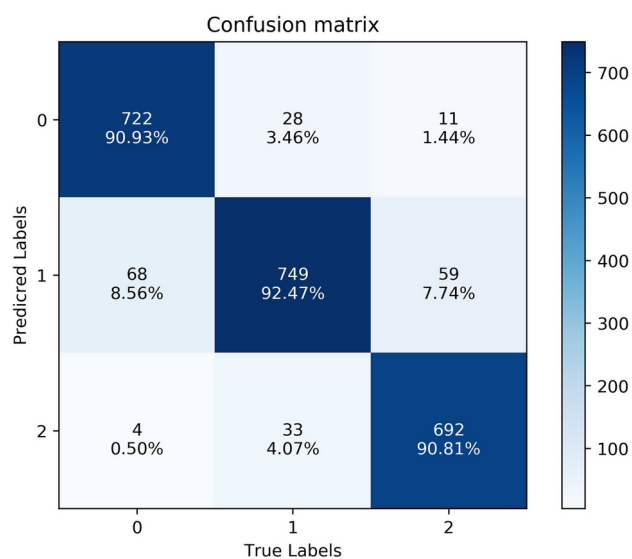


Fig. 2. Confusion matrix of the ConvNeXt model on the test set. The labels "0", "1", and "2" correspond to mild, moderate, and severe regurgitation, respectively. The confusion matrix indicates that the sensitivity for mild, moderate, and severe TR is 90.93%, 92.47%, and 90.81%, respectively. The primary prediction errors occur at adjacent severity levels; for instance, most severe labels incorrectly predicted were classified as moderate, while most mild labels incorrectly predicted were also classified as moderate.

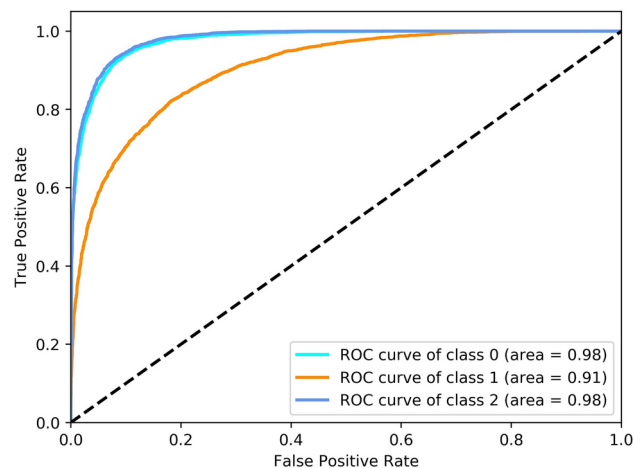


Fig. 3. ROC curves of the ConvNeXt model on the test set. The class 0 corresponds to mild regurgitation, The class 1 to moderate regurgitation, and the class 2 to severe regurgitation. The ROC curve reveals that the AUCs of the model in distinguishing mild, moderate, and severe single cardiac cycle regurgitation spectra are 0.98, 0.91, and 0.98, respectively. All AUCs exceed 0.9, demonstrating the model's efficacy in accurately identifying single cardiac cycle regurgitation spectra.

TR severity	Internal dataset				External dataset			
	Sensitivity (%)	Precision (%)	F1 Score (%)	AUC	Sensitivity (%)	Precision (%)	F1 score (%)	AUC
Mild	87.80	97.12	92.23	0.88	88.00	91.67	89.80	0.86
Moderate	86.80	84.77	85.77	0.84	85.54	86.94	86.23	0.79
Severe	92.80	86.57	89.58	0.89	91.46	87.50	89.44	0.87
Macro average	89.13	89.49	89.19	0.87	88.33	88.70	88.49	0.84

Table 4. The evaluation metrics of the Deep learning system on the internal and external datasets.

for predicting the severity of TR. To test the effectiveness of the system in assessing the severity of TR, we conducted two types of validation experiments: one utilizing an internal dataset and the other employing an external dataset.

In the internal validation phase, we selected a cohort of 1500 patients diagnosed with TR through echocardiography at our institution. The dataset consisted of 500 cases each of mild, moderate, and severe TR. By inputting the raw TR CW Doppler spectrum into the deep learning system, the prediction outcomes of TR severity for the 1500 cases in the internal validation cohort could be obtained. According to prediction results, we calculated various evaluation metrics of the deep learning system for predicting TR severity, including Accuracy, MAS, MAP, MAF1, and AUC. The internal validation results indicate that the Accuracy of the deep learning system is 89.13%. Other evaluation metrics are presented in Table 4. The confusion matrix presented in Fig. 4 elucidates the sensitivity of the deep learning system across various categories. As illustrated in Fig. 4, it is evident that the majority of prediction errors are concentrated between adjacent severity levels. Additionally, Fig. 5 presents the ROC curves of the deep learning system on the internal dataset. The ROC curves indicate that the AUCs for mild, moderate, and severe cases are 0.88, 0.84, and 0.89, respectively. The AUCs for all three categories exceed 0.8, signifying the system's excellent performance in predicting TR severity from raw CW Doppler TR spectra. The combined results presented in Table 4, Figs. 4, and 5 illustrate that the deep learning system performs excellently in assessing the severity of TR on the internal dataset.

During the external validation phase, we selected a cohort of 573 patients diagnosed with TR via echocardiography from seven external hospitals. The dataset comprised 125 cases of mild TR, 249 cases of moderate TR, and 119 cases of severe TR. Similarly, by inputting the raw TR CW Doppler spectra into the deep learning model for severity assessment, we derived the prediction outcomes for TR severity in the 573 cases of the external validation cohort, computing a series of evaluation metrics: Accuracy, MAS, MAP, MAF1, and AUC. The external validation results demonstrate the Accuracy is 88.13%, with other metrics detailed in Table 4. As demonstrated in Table 4, our system exhibited excellent performance in predicting the severity of TR in the external dataset. The confusion matrix is depicted in Fig. 6, illustrating that the majority of prediction errors are concentrated between adjacent severity levels. Figure 7 illustrates the ROC curves of the deep learning system on the external dataset. The ROC curves indicate that the AUCs for mild, moderate, and severe cases are 0.86, 0.79, and 0.87, respectively. These AUCs suggest that, except for the AUC for moderate regurgitation being slightly below 0.8, the AUCs for the other two categories are significantly above 0.8. These results indicate that despite variations in the diagnostic capabilities of cardiologists, image acquisition techniques, and equipment conditions

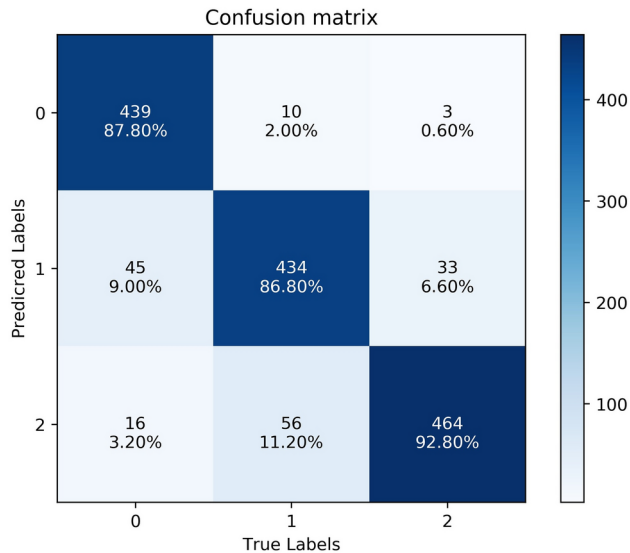


Fig. 4. Confusion matrix of the learning system on the internal dataset. The labels "0," "1," and "2" denote mild, moderate, and severe regurgitation, respectively. The confusion matrix illustrates the concordance between the TR severity predicted by the deep learning system and the clinical diagnostic results in the internal test dataset. The sensitivity for clinically diagnosed mild, moderate, and severe TR are 87.80%, 86.80%, and 92.80%, respectively. The main prediction errors primarily occur at adjacent severity levels.

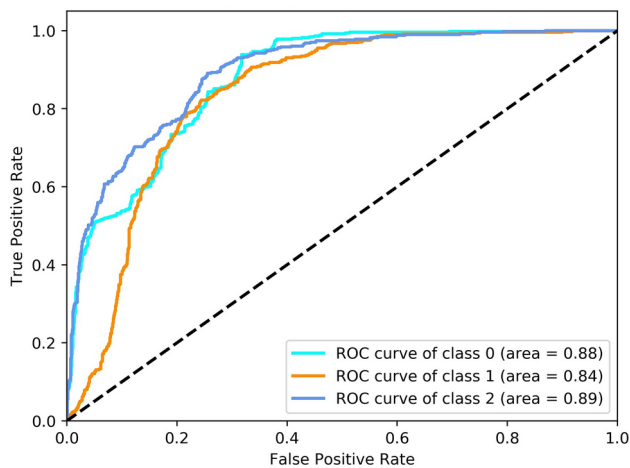


Fig. 5. ROC curves of the deep learning system on the internal dataset. The class 0 denotes mild regurgitation, the class 1 denotes moderate regurgitation, and the class 2 denotes severe regurgitation. The ROC curves indicate that the AUCs for the deep learning system in predicting mild, moderate, and severe regurgitation in the internal dataset are 0.88, 0.84, and 0.89, respectively. All AUCs exceed 0.8, underscoring the system's efficacy in internal validation experiments.

across different hospitals, our method consistently maintains high accuracy, sensitivity, and precision in TR prediction. This also demonstrates the robust generalization capability of our method.

Discussion

This study presented a novel deep learning-based approach for the intelligent assessment of TR severity utilizing the raw CW Doppler spectra. This approach was realized through an end-to-end deep learning framework integrating both segmentation and classification models. The overall efficacy of this approach was rigorously evaluated via comprehensive internal and external validation experiments. The validation outcomes indicated that this approach exhibits high overall performance on the internal and external datasets. They also suggested that this approach could serve as an intelligent assistive tool for the clinical evaluation of TR severity. However, it is important to realize that there remains room for improvement in the accuracy of severity classification, particularly in the identification of moderate regurgitation, which is notably lower than that of the other two categories. Although the discrepancy between internal and external validation results was minimal, the quantity

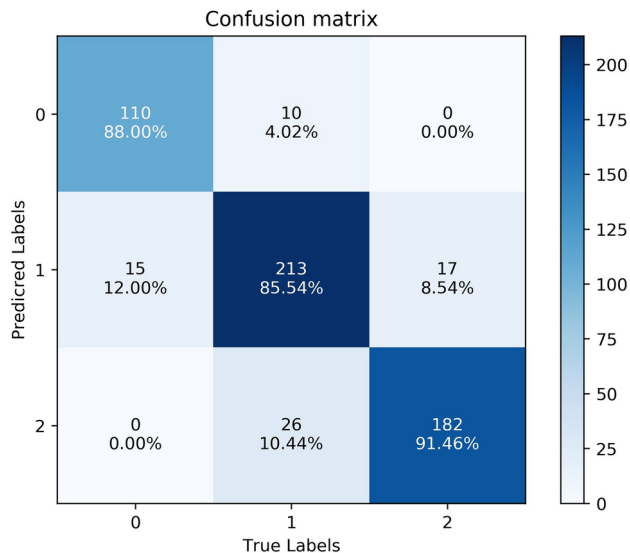


Fig. 6. Confusion matrix of the deep learning system on the external dataset. The labels "0", "1", and "2" correspond to mild, moderate, and severe regurgitation, respectively. The confusion matrix demonstrates the consistency between the predicted TR severity of the external dataset by the deep learning system and the clinical diagnostic severity. The sensitivity for mild, moderate, and severe TR diagnosed clinically is 88.00%, 85.54%, and 91.46%, respectively. The main prediction errors primarily occur at adjacent severity levels.

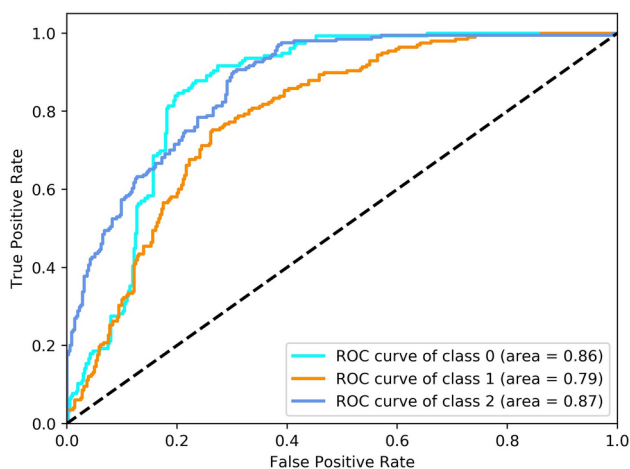


Fig. 7. ROC curves of deep learning system on the external dataset. The class 0 denotes mild regurgitation, the class 1 denotes moderate regurgitation, and the class 2 denotes severe regurgitation. The ROC curves indicates that the AUCs for the deep learning system in predicting mild, moderate, and severe regurgitation within the external dataset are 0.86, 0.79, and 0.87, respectively. Except for the AUC of 0.79 for moderate regurgitation, which is close to 0.8, the AUCs for the other two categories exceed 0.8. This further corroborates the system's efficacy in external validation experiments.

of external validation data collected in this study is less than that of the internal validation data. Therefore, future research should aim to increase the volume of external validation data. To enhance the system's generalizability, subsequent studies will incorporate data from multiple hospitals to further refine and train the deep learning model.

Traditional methods for assessing TR primarily rely on various quantitative, semi-quantitative, and qualitative indicators, such as the diameter of the inferior vena cava, regurgitant jet area, vena contracta width, and CW Doppler velocity profile image characteristics. These evaluations depend heavily on the clinical experience of cardiologist, as both the measurement of these parameters and the identification of imaging features are performed manually. Consequently, traditional methods lack objectivity, stability, and reproducibility due to human measurement errors and subjective evaluations²⁰. To meet the demand for an objective evaluation of TR, we have developed an intelligent assessment methodology for determining TR severity. Our approach introduces several innovations. Firstly, we designed a unique deep learning system comprising two models: one

for the automated segmentation of a single cardiac cycle regurgitation spectrum and another for the intelligent assessment of TR severity based on this spectrum. This is the first application of TR CW Doppler spectra for intelligent severity assessment, significantly simplifying the evaluation process. Secondly, our method provides an "end-to-end" solution that requires no manual intervention. By merely inputting the raw TR CW Doppler spectra, the assessment results are automatically generated. Thirdly, through rigorous testing and validation with both internal and external datasets, we have demonstrated the feasibility and robustness of this intelligent and objective evaluation methodology. Our methodology has proven to possess strong generalizability, ensuring reliable assessments across diverse datasets.

Although several deep learning-based techniques have been developed for evaluating valvular regurgitation via echocardiography, these techniques predominantly target the intelligent assessment of mitral and aortic valve regurgitation^{34–37}. However, compared to our method, existing approaches are more complex, demand higher standards for image data annotation in deep learning, and entail relatively intricate selection of input images in clinical applications. Our method simplifies this process by requiring only a bounding box to annotate a single cardiac cycle regurgitation image and its corresponding severity. We propose an end-to-end deep learning system that requires only raw TR CW Doppler spectra with adequate quality as input. It is worth noting that our method stipulates specific quality requirements for the CW Doppler spectra, as follows: the observation position of the regurgitation spectra should be as close as possible to the center of the regurgitation, and the direction of the Doppler sound beam should be aligned as closely as possible with the center of the regurgitation; the process of collecting the CW Doppler regurgitation spectra should minimize respiratory interference; the velocity scale in the CW Doppler images should be appropriately set. In addition, we would like to clarify that evaluating regurgitation based on single cardiac cycle regurgitation spectra may be more susceptible to interference from other factors (such as respiration, angle, etc.), which can affect the stability of the assessment results. Future research could involve selecting multiple consecutive cardiac cycle regurgitation spectra for evaluation under the premise of image quality control, which may help reduce the impact of external factors.

Meanwhile, the internal and external validation experimental results indicate that our method necessitates further enhancements in several critical areas. (1) Refinement of Classification Method: To better align with clinical diagnostic requirements, the classification methodology for assessing regurgitation severity requires further refinement. Beyond the existing categories of mild, moderate, and severe, it is more precise to incorporate additional intermediate categories, such as mild-to-moderate and moderate-to-severe, etc. This more detailed categories are expected to significantly enhance the precision of intelligent classification evaluations. (2) Inclusion of Additional Evaluation Indices: The current study is limited to the CW Doppler spectra characteristics of TR, thereby resulting in a singular evaluation index. Future research should integrate a broader range of evaluation indices to establish an intelligent assessment method based on a multi-indicator comprehensive evaluation framework. (3) Quality Control of Echocardiographic Images: The quality of echocardiographic images, which contain crucial information regarding TR, is essential for accurate diagnosis. Hence, future research should explore robust quality control methodologies for TR echocardiographic images to mitigate diagnostic errors fundamentally. Although this is a retrospective analysis of daily clinical scenarios, TR CW doppler spectrum may be influenced by multiple factors. However, we have applied strict image selection and annotation for the deep learning process. And the good results demonstrated that this method can be widely applied in our clinical practice.

Methods

Data acquisition and screening

The primary data used in this study are TR CW Doppler spectra obtained from echocardiographic examinations. We retrospectively selected echocardiographic diagnostic records from a total of 13,727 patients who were ≥ 18 years of age and diagnosed with TR. Specifically, records from 13,154 patients at our institution (Sichuan Provincial People's Hospital) constituted the internal dataset, whereas records from 573 patients at seven external hospitals comprised the external dataset. The internal dataset included patients diagnosed with TR who underwent echocardiographic examinations at our institution from January 2015 to June 2023. The external dataset consisted of patients diagnosed with TR between July 2023 and November 2023 at seven external hospitals through echocardiographic examinations. In the internal dataset, which included 13,154 patients, the raw TR CW Doppler spectra from 11,654 patients were used to develop the deep learning system, while the raw TR CW Doppler spectra from the remaining 1,500 patients were employed to validate the system's feasibility and effectiveness. In the external dataset, the raw TR CW Doppler spectra from a total of 573 patients across seven external hospitals were utilized to test the system's generalizability. To ensure the reliability and stability of the intelligent evaluation method for TR, the selected research data were meticulously reviewed by two senior cardiologists. The criteria for selected image data were as follows: (1) According to the echocardiographic diagnostic guidelines³⁸, measurement data and retained images were rigorously consistent with diagnostic conclusions; (2) TR spectra were devoid of significant respiratory or other interference; (3) The placement of the observation point and sampling line for the TR spectra was appropriate; (4) The spectra maintained a consistent velocity scale and appropriate overall gain. Furthermore, to safeguard patient privacy, all patient-related information was removed from the data. The study protocol was approved by the Ethics Review Committee of our institution (no. 2023-407) and complied with the declaration of Helsinki.

Annotation of image data

Each frame of raw TR CW Doppler spectrum encompasses multiple cardiac cycles of regurgitation spectra. In the diagnosis of TR via echocardiography, cardiologists typically evaluate the severity of regurgitation by selecting a representative single cardiac cycle regurgitation spectrum. Therefore, to achieve the intelligent assessment of TR

severity, it is crucial to develop an automatic segmentation model for single cardiac cycle regurgitation spectra, and an intelligent recognition model capable of identifying these regurgitation spectra.

To develop an automatic segmentation model for the regurgitation spectrum within a single cardiac cycle, it is essential to manually annotate the regurgitation spectrum in the raw TR CW Doppler spectra. An annotation software was employed for this purpose, and an example of the annotation is shown in Fig. 1a. A total of 15,345 single cardiac cycle regurgitation spectra from 11,654 patients' raw TR CW spectra were manually annotated. These annotations were randomly divided into training and testing sets in a 9:1 ratio, which were subsequently used to train a deep learning model for the automatic segmentation of single cardiac cycle regurgitation spectra.

Utilizing the segmentation model for single cardiac cycle regurgitation spectra, we automatically segmented and cropped the TR spectra of 11,654 patients. The cropped single cardiac cycle regurgitation spectra were screened and categorized into three severity levels: mild, moderate, and severe, as depicted in Fig. 1a. The primary criteria for these classifications adhere to echocardiographic diagnostic guidelines for TR and corresponding diagnostic outcomes. We meticulously annotated a total of 23,749 frames of single cardiac cycle regurgitation spectra, categorizing them into three severity levels: mild (7942), moderate (8129), and severe (7678). These annotated images were subsequently divided into training, validation, and test sets in a ratio of 8:1:1, forming a comprehensive dataset for the development of a deep learning model to classify TR severity based on single cardiac cycle Doppler spectra.

Training deep learning models for the segmentation of single cardiac cycle TR spectra

The automatic segmentation of single cardiac cycle regurgitation spectra is a type of object detection task. Hence, we selected multiple deep learning networks renowned for their outstanding performance in object detection, including YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, YOLOv3, YOLOv6, Faster R-CNN, and SSD, and trained them on a meticulously annotated image dataset. The performance of the automatic segmentation models was assessed using the test set. Through a comparative analysis of the test results, the model demonstrating the best overall performance was selected as the final model for automatic segmentation.

Training deep learning models for the classification of single cardiac cycle TR spectra

Classifying single cardiac cycle regurgitation spectra according to severity is essentially an image classification problem. Considering the image features of regurgitation spectra and the requirement for classification accuracy, we employed multiple deep learning networks (ConvNeXt, RepVGG, Swin Transformer, EfficientNetV2, and ResNet50) known for their robust performance in classification tasks to train models for classifying regurgitation spectra. Model performance was evaluated using the test set, and the model demonstrating superior overall performance was chosen for the classification model of single cardiac cycle TR spectra.

Establishing an end-to-end deep learning system by integrating both the segmentation and classification models

Utilizing the segmentation model described earlier, we divided a raw TR CW Doppler spectrum into multiple spectra, each representing a complete cardiac cycle spectrum. Subsequently, we can classify the severity of each cardiac cycle regurgitation spectrum using the classification model. Ideally, the severity assessment results from different cardiac cycle spectra within the same raw TR CW Doppler spectrum should be consistent. However, factors like respiration may introduce interference during spectrum acquisition, resulting in distortions of individual spectra. Consequently, the severity assessment results across different cardiac cycle spectra within the same raw TR CW Doppler spectrum may vary. To address this variability, we adopt the highest severity assessment result among multiple evaluations as the final outcome. With this final decision, we integrate the deep learning segmentation and classification models to establish an end-to-end deep learning system for TR assessment using CW Doppler spectra.

Verification of the deep learning system

To evaluate the performance of the deep learning system in assessing TR severity, we conducted validation tests utilizing both internal and external datasets. In the internal dataset test, we selected TTE examinations and diagnostic records from 1500 TR patients diagnosed at Sichuan Provincial People's Hospital. In the external data test, we collected TTE examinations and diagnostic records from 573 TR patients across seven different hospitals. The TTE examinations and diagnostic records of these TR patients were meticulously reviewed by senior cardiologists. Only TR CW Doppler spectra meeting quality requirement were utilized to predict regurgitation severity. In the detailed test experiments, CW Doppler spectra from both internal and external datasets were inputted into our end-to-end deep learning system to predict TR severity for each patient. The prediction results were subsequently compared with the echocardiographic diagnostic results reviewed by senior cardiologists, thereby rigorously evaluating the performance of the deep learning system using evaluation metrics.

Evaluation metrics

In this study, we selected the following evaluation metrics: Accuracy, Precision, Sensitivity (also known as Recall), F1 Score, Macro Average Sensitivity (MAS), Macro Average Precision (MAP), Macro Average F1 score (MAF1), mean Average Precision at 0.5 (mAP@0.5), ROC curve, AUC, and confusion matrix.

The formulas for Accuracy, Precision, Sensitivity, F1 Score, MAP, MAS, and MAF1 are as follows^{39,40}:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$Sensitivity_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$F1\ score_i = 2 \times \frac{Precision_i \times Sensitivity_i}{Precision_i + Sensitivity_i} \quad (4)$$

In Eq. (1), TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. In Eqs. (2) and (3), TP_i represents the true positives in class i , FP_i denotes the false positives in class i , and FN_i indicates the false negatives in class i . MAP , MAS , $MAF1$ are computed by taking the arithmetic mean of the metrics for individual classes.

$$MAP = \frac{1}{K} \sum_{i=1}^K Precision_i \quad (5)$$

$$MAS = \frac{1}{K} \sum_{i=1}^K Sensitivity_i \quad (6)$$

$$MAF1 = \frac{1}{K} \sum_{i=1}^K F1Score_i \quad (7)$$

Average Precision (AP) refers to the area under the precision-recall curve for a specific query or category. The mean Average Precision (mAP) is the mean of the Average Precision across all queries or categories. Assuming K queries or categories, the Average Precision for each query or category is calculated as follow:

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (8)$$

The mean Average Precision at an intersection over union (IoU) threshold of 0.5, denoted as $mAP@0.5$, is a widely used metric for performance evaluation in object detection tasks.

Data availability

The data presented in this study are available on request from the corresponding author (YW).

Code availability

The code for the automatic segmentation model of single cardiac cycle images and the classification prediction model for single cardiac cycle spectra are available upon request.

Received: 10 July 2024; Accepted: 4 November 2024

Published online: 11 November 2024

References

- Singh, J. P. et al. Prevalence and clinical determinants of mitral, tricuspid, and aortic regurgitation (the Framingham heart study). *Am. J. Cardiol.* **83**(6), 897–902 (1999).
- Nkomo, V. T. et al. Burden of valvular heart diseases: A population-based study. *Lancet.* **368**(9540), 1005–1011 (2006).
- Dong, N. et al. Advances in the treatment of heart valve disease. *J. Clin. Cardiol.* **38**(6), 429–432 (2022).
- Muraru, D. et al. Comprehensive tricuspid valve evaluation by echocardiography in the era of transcatheter therapies: A standardized approach according to the TRIAFTA study group. *J. Am. Soc. Echocardiogr.* **33**(4), 479–501. <https://doi.org/10.1016/j.echo.2019.12.002> (2020).
- Baumgartner, H. et al. Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. *J. Am. Soc. Echocardiogr.* **22**(1), 1–23 (2009).
- Anwar, A. M. et al. Assessment of normal tricuspid valve anatomy in adults by real-time three-dimensional echocardiography. *Int. J. Cardiovasc. Imaging* **23**(6), 717–724 (2007).
- Muraru, D. et al. 3-dimensional echocardiography in imaging the tricuspid valve. *JACC Cardiovasc. Imaging.* **12**(3), 500–515. <https://doi.org/10.1016/j.jcmg.2018.10.035> (2019).
- Otto, C. M. et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J. Am. Coll. Cardiol.* **77**(4), e25–e197 (2021).
- Zoghbi, W. A. et al. 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.* **30**(4), 303–371 (2017).
- Rudski, L. G. et al. Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American Society of Echocardiography. *J. Am. Soc. Echocardiogr.* **23**(7), 685–713. <https://doi.org/10.1016/j.echo.2010.05.010> (2010).
- Lancellotti, P. et al. European Association of Echocardiography recommendations for the assessment of valvular regurgitation. Part 1: aortic and pulmonary regurgitation (native valve disease). *Eur. J. Echocardiogr.* **11**(3), 223–244 (2010).
- Lancellotti, P. et al. European Association of Echocardiography recommendations for the assessment of valvular regurgitation Part 2: mitral and TR (native valve disease). *Eur. J. Echocardiogr.* **11**(4), 307–332 (2010).

13. Lancellotti, P. et al. Recommendations for the echocardiographic assessment of native valvular regurgitation: an executive summary from the European Association of Cardiovascular Imaging. *Eur. Heart J. Cardiovasc. Imaging* **14**(7), 611–644. <https://doi.org/10.1093/ehjci/jet105> (2013).
14. Wifstad, S. V. et al. Quantifying valve regurgitation using 3-D Doppler ultrasound images and deep learning. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **69**(12), 3317–3326 (2022).
15. Addetia, K. et al. 3-Dimensional echocardiographic analysis of the tricuspid annulus provides new insights into tricuspid valve geometry and dynamics. *JACC Cardiovasc. Imaging.* **12**(3), 401–412. <https://doi.org/10.1016/j.jcmg.2017.08.022> (2019).
16. Thomas, N. et al. Intraobserver variability in grading severity of repeated identical cases of mitral regurgitation. *Am. Heart J.* **156**(6), 1089–1094 (2008).
17. Biner, S. et al. Reproducibility of proximal isovelocity surface area, vena contracta, and regurgitant jet area for assessment of mitral regurgitation severity. *JACC Cardiovasc. Imag.* **3**(3), 235–243 (2010).
18. Buck, T. et al. Effect of dynamic flow rate and orifice area on mitral regurgitant stroke volume quantification using the proximal isovelocity surface area method. *J. Am. Coll. Cardiol.* **52**(9), 767–778 (2008).
19. Bargiggia, G. S. et al. A new method for quantitation of mitral regurgitation based on color flow Doppler imaging of flow convergence proximal to regurgitant orifice. *Circulation.* **84**(4), 9–1481 (1991).
20. Hahn, R. T. Imaging assessment of TR severity. *JACC Cardiovasc. Imaging* **12**(3), 469–490 (2019).
21. Dey, D. et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **76**(22), 2668–2681 (2020).
22. Ghorbani, A. et al. Trivial mitral valve regurgitation detection using convolutional neural networks. *J. Med. Syst.* **44**(2), 44 (2020).
23. Kamphuis, V. P. et al. Deep learning-based quantification of mitral regurgitation using cardiac cine MRI. *Radiol. Cardiothorac. Imaging* **2**(2), e190005 (2020).
24. Singh, A. et al. Artificial intelligence in valvular heart disease: current status, applications, and future prospects. *Front. Cardiovasc. Med.* **8**, 77 (2021).
25. Zhang, Q. et al. Automatic assessment of mitral regurgitation severity using the mask R-CNN algorithm with color Doppler echocardiography images. *Comput. Math. Methods Med.* **13**(2021), 2602688. <https://doi.org/10.1155/2021/2602688> (2021).
26. Huang, J., Huang, A., Xu, R., Wu, M., Wang, P., Wang, Q. Automatic segmentation and assessment of valvular regurgitations with color Doppler echocardiography images: A VABC-UNet-based framework. *Bioengineering* (2023)
27. Jocher, G., Stoken, A., Borovec, J. et al. ultralytics/yolov5: v5. 0-YOLOv5-P6 1280 models, AWS, Supervise. ly and YouTube integrations. *Zenodo.* (2021)
28. Jocher, G., Chaurasia, A., Stoken, A. et al. ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations. *Zenodo* (2022)
29. Liu, Z., Mao, H., Wu, C. Y. et al. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11976–11986 (2022)
30. Ding, X., Zhang, X., Ma, N. et al. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 13733–13742 (2021).
31. Liang, J., Cao, J., Sun, G., et al. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 1833–1844 (2021).
32. Tan, M., Le, Q. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning* 10096–10106 (PMLR, 2021)
33. Theekedath, D. & Sedamkar, R. R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **1**(2), 79 (2020).
34. Atika, L. et al. Image segmentation for mitral regurgitation with convolutional neural network based on UNet, Resnet, Vnet, FractalNet and SegNet: A preliminary study. *Big Data Cogn. Comput.* **6**(4), 141 (2022).
35. Vrudhula, A., Duffy, G., Vukadinovic, M., et al. High throughput deep learning detection of mitral regurgitation. medRxiv. 2024.02.08.24302547 (2024)
36. Edwards, L. A. et al. Machine learning for pediatric echocardiographic mitral regurgitation detection. *J. Am. Soc. Echocardiogr.* **36**(1), 96–104 (2023).
37. Brown, K. et al. Using artificial intelligence for rheumatic heart disease detection by echocardiography: Focus on mitral regurgitation. *J. Am. Heart Assoc.* **13**(2), e031257–e031257 (2024).
38. Zoghbi, W. A. et al. 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.* **30**(4), 303–371. <https://doi.org/10.1016/j.echo.2017.01.007> (2017).
39. Lee, K. W., Lee, H. J., Hu, H. & Kim, H. J. Analysis of facial ultrasonography images based on deep learning. *Sci. Rep.* **12**(1), 16480. <https://doi.org/10.1038/s41598-022-20969-z> (2022).
40. Shah, A. A. et al. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Sci. Rep.* **13**, 2987. <https://doi.org/10.1038/s41598-023-29656-z> (2023).

Author contributions

S.X. and H.L.: research concept, study design, training models, data analysis and interpretation, writing of the manuscript. L.S., J.S., J.M., D.H., M.Z., H.L. and Y.L.: data collection, and illustration, L.Y.: manuscript draft reviewing/editing, Q.S. and Y.W. data collection and illustration, data analysis and interpretation, manuscript draft reviewing/editing, manuscript writing supervision. All authors provided critical feedback, reviewed the paper, and approved the manuscript.

Funding

This research was supported by Sichuan Provincial Science and Technology Plan Project (2023YFQ0006).

Declarations

Competing interests

The authors declare no competing interests.

Statement of ethics

Our study is a retrospective study. Our research protocol was reviewed and approved by the Ethics Review Committee of our institution (no. 2023-407). Since the data collected in this study consist of images and diagnostic reports that are routinely stored during echocardiography examinations, the Ethics Committee also approved the waiver of informed consent for this study.

Additional information

Correspondence and requests for materials should be addressed to L.Y., Q.S. or Y.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024