Ultrasound in Medicine & Biology 49 (2023) 2247−2255

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

journal homepage: www.elsevier.com/locate/ultrasmedbio

Original Contribution

Automatic Myocardial Contrast Echocardiography Image Quality Assessment Using Deep Learning: Impact on Myocardial Perfusion Evaluation

Mingqi Li [a], Dewen Zeng [b], Hongwen Fei [c], Hongning Song [a], Jinling Chen [a], Sheng Cao [a], Bo Hu [a], Yanxiang Zhou [a], Yuxin Guo [a], Xiaowei Xu [d], Kui Huang [a], Ji Zhang [a], Qing Zhou [a,*]

[a] Department of Ultrasound Imaging, Renmin Hospital of Wuhan University, Wuhan, China
[b] Department of Computer Science and Engineering, University of Notre Dame, South Bend, IN, USA
[c] Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China
[d] Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China

A R T I C L E I N F O

Keywords:
Myocardial contrast echocardiography
Image quality assessment
Artificial intelligence
Deep learning
Myocardial perfusion

Objective: The image quality of myocardial contrast echocardiography (MCE) is critical for precise myocardial perfusion evaluation but challenging for echocardiographers. Differences in quality may lead to diagnostic heterogeneity. This study was aimed at achieving automatic MCE image quality assessment using a deep neural network (DNN) and investigating its impact on myocardial perfusion evaluation.
Methods: The Resnet-18 model was used for training and testing on internal and external data sets. Quality assessment involved three aspects: left ventricular opacification (LVO), shadowing, and flash adequacy; the quality score was calculated based on image quality. This study explored the impact of the DNN-based quality score on perfusion evaluation (normal, delay or obstruction) by echocardiographers (two seniors, one junior and one novice). Additionally, the effect of the score difference between re-scans on perfusion evaluation was investigated.
Results: The time cost for DNN prediction was 0.045 s/frame. In internal validation and external testing, the DNN achieved F1 and macro F1 scores >90% for quality assessment and had high intraclass correlation coefficients (0.954 and 0.892, respectively) in sequence quality scores. The proportion of segments deemed uninterpretable increased as the DNN-based quality score decreased. The agreement of perfusion assessment between one senior and others decreased as the quality score decreased. And the greater the score difference between the re-scans, the lower was the agreement on perfusion assessment by the same echocardiographer.
Conclusion: This study determined the effectiveness of DNN for real-time automatic MCE quality assessment. It has the potential to reduce the variability in perfusion evaluation among echocardiographers.

Introduction

The clinical application of ultrasonic enhancing agents (UEAs) in echocardiography is becoming increasingly widespread. Myocardial contrast echocardiography (MCE) offers a distinct advantage in the evaluation of myocardial perfusion by "flash" (high-mechanical-index [MI] impulses) technology [1]. However, the real-time image quality control and image interpretation of MCE are challenging and highly dependent on experience.

Myocardial contrast echocardiography requires high and special image quality because myocardial perfusion analysis is based on the change in grayscale of myocardial segments over time after "flash." Artifacts in images may result in inaccurate presentation. And the differences in image quality may lead to diagnostic heterogeneity [2,3]. The European Society of Cardiology emphasizes that some perfusion defects in MCE may be caused by image quality issues [4]. So, real-time image quality assessment is a prerequisite for accurate perfusion analysis.

The imaging mechanism underlying MCE is elimination of myocardial tissue signals and retention of microbubble signals for imaging [5]. So there are more floating speckles than in conventional echocardiography, and the grayscale displayed is influenced by factors such as the "flash" settings, the scanning technique of the sonographer, the injection speed of the UEAs and the rate of systemic circulation [6]. However, MCE image quality control currently poses a significant challenge for beginners or junior echocardiographers. Even senior echocardiographers may miss image quality issues because of lapses in attention or work fatigue, resulting in diagnostic errors.

Artificial intelligence (AI) has great potential to address these issues in image quality assessment. AI involves the use of algorithms to

* Corresponding author. Department of Ultrasound Imaging, Renmin Hospital of Wuhan University, No. 238 Jiefang Road, Wuchang District, Wuhan 430060, Hubei Province, China.
E-mail address: qingzhou.wh.edu@hotmail.com (Q. Zhou).

https://doi.org/10.1016/j.ultrasmedbio.2023.07.002
Received 14 May 2023; Revised 12 June 2023; Accepted 3 July 2023

simulate human logic or knowledge frameworks with the aim of achieving human activities. Machine learning is a crucial approach, in which algorithms are designed to learn how to execute tasks by analyzing expert-crafted data feature descriptions [7]. Deep learning is a type of machine learning that uses convolutional neural networks to learn directly from vast amounts of data, enabling it to perform targeted tasks without requiring manual feature engineering [8]. Currently, the application of deep neural network (DNNs) in image quality assessment in MCE is almost blank.

Therefore, the purpose of this study was to use ResNet-18 to achieve automatic MCE image quality assessment from three aspects—left ventricular opacification (LVO), acoustic shadowing and flash adequacy—and simultaneously output quality score. Furthermore, we aimed to investigate the impact of DNN-based quality score on the evaluation of myocardial perfusion by different echocardiographers.

Methods

Imaging data set

This retrospective study was performed and approved by the ethics committee at Renmin Hospital of Wuhan University (ID: WDRY2022-K270). From June 2021 to June 2022, a total of 1136 MCE image sequences were retrospectively included as the image data set in this study, containing 354 apical four-chamber view (A4C) sequences, 371 apical two-chamber view (A2C) sequences and 411 apical three-chamber view (A3C) sequences. To maximize the efficiency of image data set utilization, we set a frame interval of 1 frame and collected a maximum of 100 frames of images per sequence after the flash. The image sequences were divided into a training set (n = 569 sequences, 56,229 frames) and an internal validation set (n = 567 sequences, 55,273 frames). To investigate the generalizability of the proposed DNN, an external test set (191 sequences, 16,972 frames) was gathered that was referred for MCE from an independent institution (Guangdong Provincial People's Hospital). Figure 1 is the study flowchart.

Contrast echocardiography

Myocardial contrast echocardiography was performed on a Philips 7C (Philips Medical Systems, Best, Netherlands) equipped with a contrast-specific multipulse amplitude modulation imaging algorithm and a broadband transducer S5-1. The UEA SonoVue (Bracco Research SA, Geneva, Switzerland) powder was diluted to 15 mL with normal saline (2.5 mL of the initial solution + 12.5 mL normal saline). The intravenous continuous infusion rate was adjusted based on the echocardiographer's experience. Real-time contrast imaging was performed using a low-MI setting of 0.15−0.19 with a frame rate of 20−30 Hz. A transient high-MI flash (10−15 frames) with a setting of 1.20−1.30 was employed to clear myocardial microbubbles. To minimize background signals from the myocardium or blood, time-gain compensation and gain (60%−70%) were adjusted prior to UEA infusion. During UEA infusion, digital captures of the A2C, A3C and A4C views were taken with 13 to 15 cardiac cycles. All settings were kept unchanged throughout the examination. In this uncontrolled clinical practice, echocardiographers were allowed to repeat image acquisition in the same chamber view or not for any acceptable reasons.

Image quality assessment of MCE

The image quality assessment labels were initially assigned by a senior echocardiographer with 5 years of experience in MCE. Subsequently, another senior echocardiographer reviewed the initial labels for validation. In case of any disagreements, the final labeling was determined through discussions or voting involving a third senior echocardiographer. The labels and original images were then used for neural network supervised learning and predictive model validation and testing. Quality assessment consisted of three aspects, and the quality score was calculated on the basis of image quality (Fig. S1, online only) [5,6]:

1. Left ventricular opacification: Optimal (optimal LV cavity contrast from the apex to the mitral annular plane, 1 point); low (insufficient

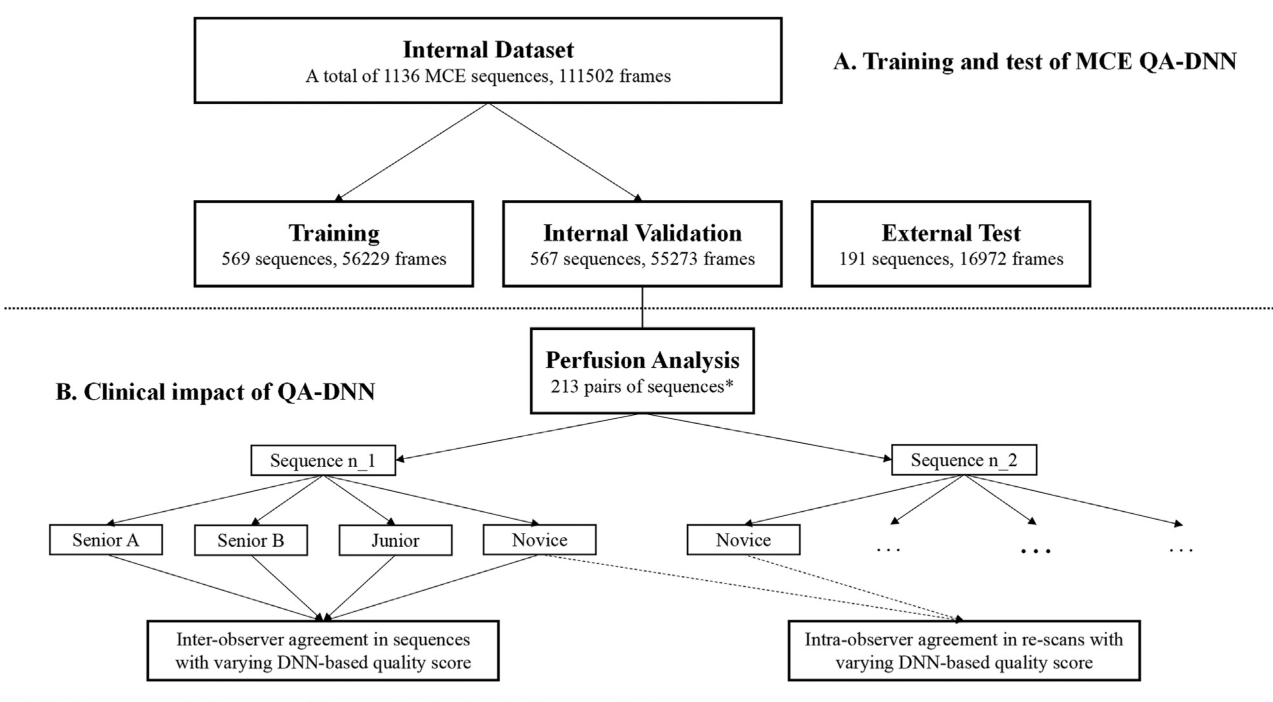

Figure 1. Study flowchart. DNN, deep neural network; MCE, myocardial contrast echocardiography; QA, quality assessment.

LV cavity contrast with swirling, 0 points); high (excessive LV cavity contrast with near-field microbubble signals overflowing to the myocardium, or/and with attenuation and shadowing at the mid- or basal regions of the LV cavity, 0 points).

2. Acoustic shadowing: No shadowing (1 point); presence of shadowing (the myocardium is obscured by the rib/lung, with the epicardium or endocardium invisible, 0 points).
3. Flash: Optimal (complete clearance of myocardial contrast signals without excessive LV cavity microbubble destruction, 1 point); insufficient (incomplete clearance of myocardial contrast signals, 0 points); excessive (no LV cavity contrast signals because of complete microbubble destruction, 0 points).

In this study, the echocardiographer labeled the classifications of LVO and shadowing for each frame. In general, evaluating the adequacy of flash requires only the first frame after flash. However, to increase the training sample size for the "flash" model, we assumed that the first 5 frames after flash in each sequence in the training set were labeled for adequacy. In the internal validation and external test sets, only the first frame after flash in each sequence was predicted for “flash” adequacy.

Deep neural network

We trained three neural networks for quality assessment of LVO (optimal, low, high), shadowing and flash (optimal, excessive, insufficient).

This study used the Resnet-18 convolutional neural network as the model architecture, as illustrated as Figure 2, which includes specific skip connections to help the network train more effectively. The study was divided into three tasks, each training a different quality assessment network. The first task was to determine whether the flash was qualified. The neural network's input was the first 5 frames after the flash, and the output was one of the flash quality categories. The second and third tasks were to classify the types of LVO and shadowing of all frames after the flash. The neural network's input was every single frame image after the flash, and the output is one of the predetermined image quality categories. All tasks use the same neural network architecture.

After the image input enters the first convolutional layer, it passes through four residual blocks, each containing two convolutional layers. The output of the last residual block is connected to a max pooling layer, a fully connected layer and a softmax layer to obtain the probability of each category. The category with the highest probability is the final prediction of the neural network. Each convolutional layer in the neural network is followed by a dropout layer, batch normalization layer and ReLU activation layer to help the neural network converge and optimize better.

The cross-entropy loss function was used for training, with the Adam optimizer and an initial learning rate of 0.001. The entire network was trained for 200 epochs, with training and validation performed in each epoch. The final network model was selected as the one with the highest accuracy on the validation set during the training process.

Myocardial perfusion analysis

In the internal validation set, there were 213 pairs of sequences (repeated scans in same chamber view for the same patient) with potentially different image qualities. To assess the clinical implications of the MCE image quality assessment DNN, we initially randomized the image sequence. Subsequently, a novice, a junior and two senior echocardiographers, blinded to all patient information and DNN-based quality score, independently performed perfusion analyses on the MCE image sequence.

The myocardial perfusion (MVP) evaluation was conducted using the following scoring system [9]: 1 = normal (time of replenishment completion: <4 s), 2 = delayed MVP (time of replenishment completion: 4−8 s), 3 = microvascular obstruction (MVO) (persistent perfusion defect). Uninterpretable segments or low-confidence segments in interpretation resulting from image quality issues were labeled as "NA" based on the echocardiographers' own judgment.

Clinical impact of DNN-based quality assessment

According to the DNN image quality scores, all sequences were classified into the high-quality group (≤3 and >2), medium-quality group (≤2 and >1) and low-quality group (≤ 1). Then, the association between image quality and the proportion of "NA" segments was explored, as was the effect of image quality on the inter-observer variability of perfusion evaluations among novice, junior, senior B and senior A.

By use of the difference in DNN-based quality scores of paired sequences, the sequences were equally divided into a high-difference group and a low-difference group. The effect of image quality difference on the intra-observer variability of perfusion evaluations among echocardiographers was then explored in paired sequences.

Statistical analysis

Data were analyzed using SPSS 25.0 (IBM, Armonk, NY, USA). The performance of the DNN was evaluated using recall, precision, F1 score or macro F1 score, using the expert-labeled image quality issues as the

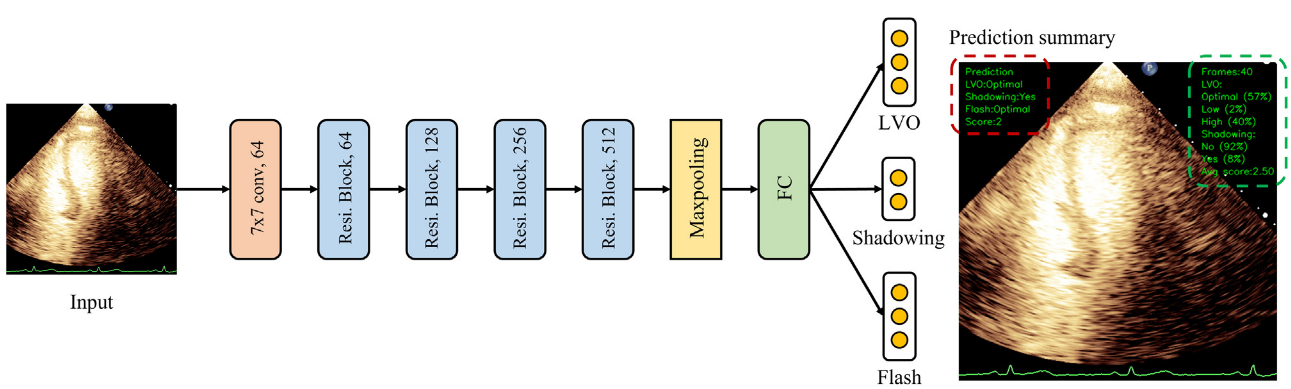

Figure 2. ResNet18-based myocardial contrast echocardiography image quality assessment framework. This is a simplified version of the ResNet-18 architecture. In the prediction summary, the *red box* indicates the deep neural network's prediction and score for the quality of the current frame. The *green box* displays a summary of the image quality issues for all frames up to the current frame, along with their average score. 7 × 7 conv 64, a convolutional layer with a kernel size of 7 × 7, which takes an input image and produces a feature map with 64 channels; FC, fully connected layers to obtain a set of probabilities for each category; QA, quality assessment; Resi Block, a residual block consisting of two convolutional layers and a skip connection.

ground truth. Agreement between DNN-based quality score and human visual score was evaluated using the intraclass correlation coefficient (ICC). The impact of the DNN-based quality score on perfusion evaluation was evaluated using the macro F1 score and accuracy. Categorical variables were analyzed using the χ^2-test. p Values <0.05 were considered to indicate statistical significance.

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$\text{Accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}}$$

For dichotomous prediction,

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

For trichotomous prediction,

$$\text{Precision}_{ma} = \frac{\text{precision}_1 + \text{precision}_2 + \text{precision}_3}{3}$$

$$\text{Recall}_{ma} = \frac{\text{recall}_1 + \text{recall}_2 + \text{recall}_3}{3}$$

$$\text{Macro F1 score} = 2 \times \frac{\text{precision}_{ma} \times \text{recall}_{ma}}{\text{precision}_{ma} + \text{recall}_{ma}}$$

Results

Data set

The distribution of various image quality issues in the data set is described in Table 1.

Predictive performance of DNN in image quality issues

The average latency (time cost) of DNN prediction was 0.045 s/frame. The image quality evaluation and quality scores were displayed on each frame (Fig. 3; Video S1, online only).

As outlined in Table 2, the proposed DNN achieved automatic image quality assessment with a good F1 or macro F1 score in detecting LVO (internal: 97.71%, external: 95.39%), acoustic shadowing (internal: 93.81%, external: 92.59%) and flash adequacy (internal: 96.38%, external: 91.61%).

Among all frames, the confusion matrix (Fig. S2, online only) provides a detailed display of the classification of the DNN in the internal validation and external test. In the internal validation, the recalls of optimal, low and high LVO were 99.15%, 97.51% and 95.83%, respectively; 95.22% of frames with shadowing were detected by DNN. The recalls of optimal, excessive and insufficient flash were 94.88%, 93.27% and 91.79%, respectively. During external testing, the recalls for various image quality issues ranged from 85.92% to 100.00%.

DNN-based quality scoring for sequences

The ICCs of the DNN-based quality score and human visual score on sequences were 0.954 (95% confidence interval [CI]: 0.946−0.961) and 0.892 (95% CI: 0.858−0.918) in the internal validation and external test, respectively. As illustrated in the Bland−Altman plot (Fig. 4), in the internal validation and external test, the mean differences were 0.005 and 0.054, respectively.

Impact of DNN-based quality score on myocardial perfusion analysis

Among all echocardiographers, as the DNN-based quality score decreased, the proportion of “NA” segments gradually increased (Fig. 5). For instance, in senior A, for high (n = 108) versus medium (n = 264) versus low quality (n = 54), the proportions of “NA” segments were 2.65% versus 14.72% versus 51.06%, respectively (p < 0.001). Among all sequences, the proportions of “NA” segments of novice and junior were the lowest and highest, respectively (novice *vs.* junior *vs.* senior A *vs.* senior B: 4.90% *vs.* 22.07% *vs.* 16.26% *vs.* 12.54%, respectively, $p < 0.001$).

Table 3 and Figure 6 present the inter-observer agreement among echocardiographers with different levels of experience on perfusion evaluation in sequences with varying image quality. Taking senior A as the reference, the macro F1 scores of senior B and junior decreased with decreasing image quality. For instance, for senior B, the F1 scores were 83.34%, 74.27% and 66.73%, in the sequences with high, medium and low quality, respectively. However, no such trend is observed in the macro F1 scores of the novice (which were below 50.00% at all image quality levels). In addition, the accuracies of senior B, junior and novice all decreased with decreasing image quality.

Table 4 and Figure 7 illustrate the agreement of perfusion evaluation by each echocardiographer on re-scan sequences. With the perfusion assessment on the first acquired sequence as reference, the macro F1 scores and accuracies of the low-difference group (n = 106 pair sequences) are higher than those of the high-difference group (n = 107 pair sequences). For example, in senior A, the macro F1 scores were 79.72% and 60.85%, and the accuracies were 84.77% and 69.83% in the low-

Table 1
Data set for myocardial contrast echocardiography quality assessment deep neural network

	Training	Internal validation	External test
Number of sequences (frames)	569 (56,229)	567 (55,273)	191 (16,972)
Flash			
Optimal	1441 (50.65%)[a]	173 (30.51%)	86 (45.03%)
Excessive	927 (32.58%)	383 (67.55%)	57 (29.84%)
Insufficient	477 (16.77%)	11 (1.94%)	48 (25.13%)
Left ventricular opacification			
Optimal	41,382 (73.60%)	40,684 (73.61%)	10,568 (62.27%)
Low	9660 (17.18%)	10634 (19.24%)	980 (5.77%)
High	5187 (9.22%)	3955 (7.16%)	5424 (31.96%)
Shadowing			
No	43,997 (78.25%)	42,280 (76.49%)	11,757 (69.27%)
Yes	12232 (21.75%)	12993 (23.51%)	5215 (30.73%)

[a] Number of frames (%). To increase the size of the training data set of “flash,” the "flash" data set was constructed by including the first five frames of each sequence as the training set and using the first frame of each sequence as the test set.

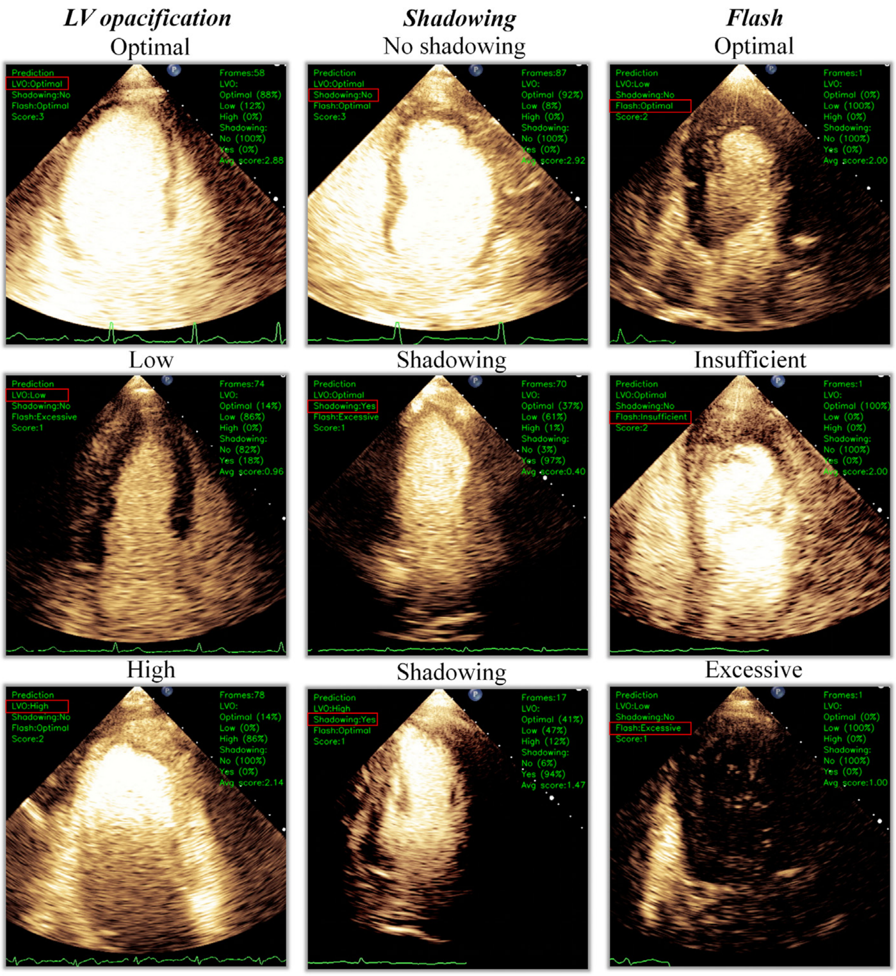

Figure 3. Deep neural network prediction for various image quality issues. LV, left ventral.

Table 2
Predictive performance of myocardial contrast echocardiography quality assessment deep neural network

	Recall	Precision	F1 score
Internal validation			
Flash			96.38%
Optimal	93.06%	92.00%	
Excessive	96.34%	96.85%	
Insufficient	100.00%	100.00%	
LVO			97.71%
Optimal	99.15%	98.98%	
Low	97.51%	97.67%	
High	95.83%	97.13%	
Shadowing			93.81%
No shadowing	98.54%	97.73%	
Shadowing	92.54%	95.12%	
External test			
Flash			91.61%
Optimal	87.21%	92.59%	
Excessive	89.47%	96.23%	
Insufficient	100.00%	84.21%	
LVO			95.39%
Optimal	98.00%	98.85%	
Low	85.92%	97.79%	
High	97.79%	94.14%	
Shadowing			92.59%
No shadowing	95.07%	98.07%	
Shadowing	95.78%	89.60%	

LVO, left ventricular opacification.

and high-difference groups, respectively. See Tables S1 and S2 (online only) for confusion matrixes of perfusion assessment by each echocardiographer.

Discussion

This study has proposed a deep learning-based framework for MCE image quality assessment that enables automatic evaluation of MCE images from the perspectives of LVO, shadowing and flash, and simultaneously outputs image quality scores. In terms of perfusion assessment, this study found that (i) the proportion of ungradable segments increased with decreasing DNN-based quality scores; (ii) the inter-observer agreement decreased with decreasing DNN-based quality scores; and (iii) the greater the difference in DNN-based quality scores between the re-scan sequences, the lower was the agreement on perfusion assessment by the same echocardiographer. Therefore, application of this real-time feedback DNN to MCE image quality control in clinical practice has the potential to improve the feasibility of perfusion assessment and reduce diagnostic heterogeneity among echocardiographers.

Impact of DNN-based MCE image quality assessment on image acquisition

Myocardial contrast echocardiography examinations require real-time adjustment of the injection rate of contrast agents, probe angle (even if the endocardial and chamber boundaries are clearly displayed, shadowing caused by ribs or lung gas can lead to inaccurate myocardial grayscale) and flash setting, which can be challenging for inexperienced echocardiographers.

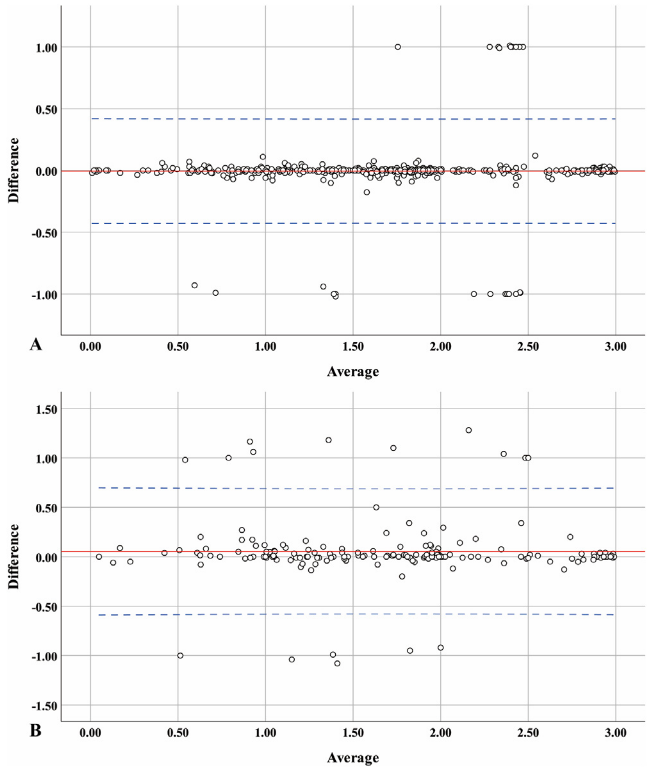

Figure 4. Bland−Altman plot comparing deep neural network−based quality scores with expert-based quality scores.

This study achieved near-expert-level automated and real-time image quality assessment through deep learning. Its fast image quality classification is expected to provide real-time guidance during image acquisition to assist inexperienced echocardiographers in making appropriate adjustments, and to aid experienced echocardiographers in avoiding issues of attentional bias. It will result in the acquisition of qualified MCE images for perfusion analysis, thereby effectively avoiding problems caused by missing data resulting from unqualified images. Additionally, in offline analysis, echocardiographers may need to select the sequence for analysis from several re-scans. The proposed DNN can be used as a tool for fast screening of high-quality images, ensuring accurate image analysis.

Another advantage of artificial intelligence is its ability to perform high-intensity tasks without getting tired, bored or distracted, which ensures consistent and accurate results. For example, sometimes shadowing occurs only in a few consecutive frames, which may be difficult for the human eye to detect. However, if it occurs at the end of systole, it will have an adverse impact on quantitative analysis. The DNN can independently detect and record image quality issues for each frame, effectively solving this problem.

Impact of MCE image quality assessment DNN on perfusion evaluation

In addition to improving boundary display of anatomical structures in echocardiography, MCE is more important for the evaluation of myocardial perfusion. MCE has important value in diagnosing microcirculation dysfunction in ST-segment elevation myocardial infection after percutaneous coronary intervention, as well as predicting prognosis [9,10]. If there are issues with MCE image quality, grayscale pixels in the image may be distorted or artifacts may appear. The change in image grayscale over time is the basis for qualitative and quantitative perfusion analysis. Therefore, image quality control of MCE is crucial for accurate evaluation of myocardial perfusion, which involves mainly LVO, unreal segment contrast and machine setting.

Excessive LVO caused by rapid injection may result in acoustic attenuation in the far field, leading to false-positive diagnosis of MVO in the basal myocardium. At the same time, it may also manifest as excessive contrast in the near field. The excessive reflection of microbubble signals in the LV cavity may mask myocardial perfusion abnormalities in the apex myocardium. Insufficient LVO indicates inadequate microbubble

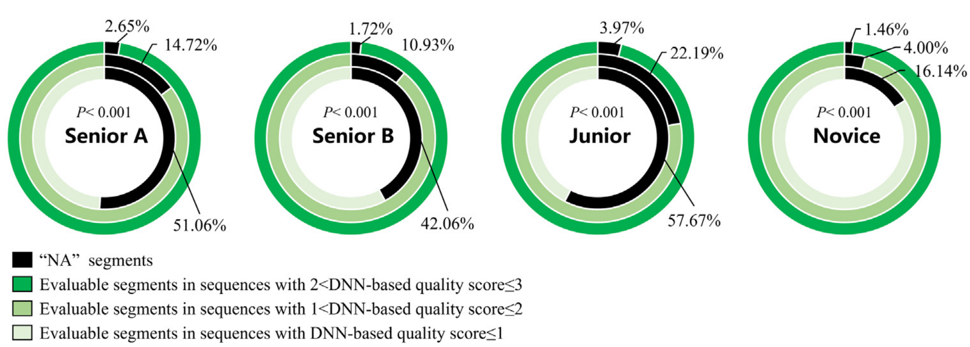

Figure 5. Proportion of uninterpretable segments among different DNN-based quality scores and echocardiographers with different levels of experience. DNN, deep neural network; "NA" segments, uninterpretable segments or low-confidence segments in interpretation because of image quality issues.

Table 3
Inter-observer agreement among echocardiographers on perfusion evaluation in sequences with varying deep neural network−based quality scores

	2 < quality score ≤ 3	1 < quality score ≤ 2	Quality score ≤ 1
Number of sequences	108	264	54
Macro F1 score			
Senior A *vs.* B	83.34%	74.27%	66.73%
Senior A *vs.* junior	61.81%	54.60%	46.59%
Senior A *vs.* novice	33.13%	44.99%	36.31%
Accuracy			
Senior A *vs.* B	90.08%	81.01%	76.46%
Senior A *vs.* junior	77.65%	62.61%	62.17%
Senior A *vs.* novice	60.98%	55.92%	35.71%

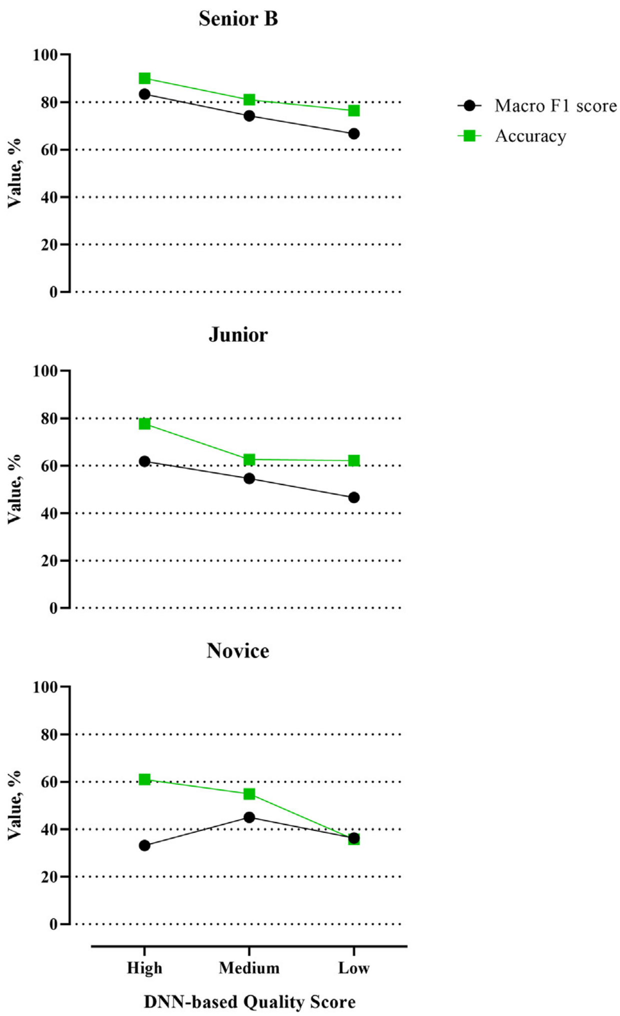

Figure 6. Association between diagnostic agreement and DNN-based quality score, with the perfusion evaluation of senior A as reference. High DNN-based quality score (n = 108): ≤3 and >2; medium DNN-based quality score (n = 264): ≤2 and >1; low DNN-based quality score (n = 54): ≤1. DNN, deep neural network.

concentration and may lead to false-positive diagnosis of all segments. When there is shadowing caused by ribs or lung, some segments cannot be evaluated for perfusion, especially in the apical two-chamber view. When excessive flash occurs, microbubbles in the LV cavity are completely cleared, and replenishment of microcirculation requires a longer time, which may lead to misdiagnosis of delayed MVP.

Table 4
Agreement of perfusion diagnosis by each echocardiographer on repeatedly acquired sequences with deep neural network−based low and high score difference

	Low-difference group	High-difference group
Number of pair sequences	106	107
Macro F1 score		
Senior A	79.72%	60.85%
Senior B	79.51%	48.42%
Junior	59.74%	49.15%
Novice	66.04%	57.96%
Accuracy		
Senior A	84.77%	69.83%
Senior B	85.85%	65.82%
Junior	71.43%	61.42%
Novice	76.02%	71.96%

Conversely, when insufficient flash occurs, microbubbles in the myocardium are not completely cleared, and replenishment to plateau may occur in a shorter time, which may result in missed diagnosis of delayed MVP.

The presence of these artifacts during perfusion analysis is deemed unacceptable in clinical practice, and adjustments should be made. For example, when attenuation occurs, it is essential to promptly reduce the injection rate or the dose of UEAs. In the presence of shadowing artifacts, it is imperative to reposition or adjust the probe angle to optimize the visualization of both sides of the myocardium, ensuring clear depiction. When dealing with patients who are obese or have chronic obstructive pulmonary disease, a practical approach involves acquiring two sequences to obtain perfusion information for each side of the myocardium separately. When encountering excessive flash, it is necessary to lower the MI of the flash or decrease the number of flash frames.

The unaddressed image quality issues may serve as a contributing factor to the diagnostic heterogeneity among echocardiographers in this study. In terms of agreement between two senior echocardiographers, the macro F1 score and accuracy of the high-quality group predicted by DNN were the highest among the three quality groups, indicating that selecting high-quality sequences can bring low diagnostic variability. This can improve the physician's diagnostic confidence and enhance the reliability of MCE examination in clinical practice. For the junior echocardiographer, providing high-quality sequences predicted by DNN can improve diagnostic yield, using senior A as reference. In the experiment on perfusion evaluation of re-scans, the greater the quality difference predicted by DNN of the two image sequences, the lower was the Macro F1 score and accuracy within observers, indicating that the image quality variability of the same patient can affect perfusion evaluation by the same echocardiographer, and these differences in diagnosis were not due to the patients' conditions.

In addition, novice physicians without MCE examination experience may have difficulty judging whether there were quality issues with the image. Therefore, in this study, the proportion of "NA" segments for the novice echocardiographer was significantly lower than that for other levels of echocardiographer, as they might have evaluated certain "NA" segments for perfusion. However, similar to other echocardiographers, the proportion of "NA" segments was negatively correlated with the predicted image quality by DNN.

In terms of quantitative analysis, attenuation induced by high LVO may lead to a decrease in A value (plateau contrast intensity) at the basal segments, while causing an increase in A value at the apical segments. Low LVO may result in a decrease in A value across all segments. Segments in shadowing may yield lower A values or may result in unsuccessful parameter fitting. Excessive flash may contribute to lower β values (mean microbubble velocity), whereas insufficient flash may lead to higher β values.

These variations in A or β values ultimately have an impact on the calculated myocardial blood flow. However, the specific effects of image quality on quantitative analysis require further in-depth exploration in future research.

Current era of deep learning on echocardiographic image quality assessment

Conventional echocardiographic image quality evaluation relies on anatomical structure visibility and ultrasound view standards. Several studies have employed deep learning to score image quality on conventional echocardiography, but only at an overall level [11−13]. A more practical approach would identify specific issues to guide ultrasound physicians [14]. Recent advancements in this field include a real-time guidance algorithm that scores image quality and suggests specific adjustments [15]. To develop such software for MCE, labels during training should also include injection adjustments and setting adjustments. Nonetheless, exploring the

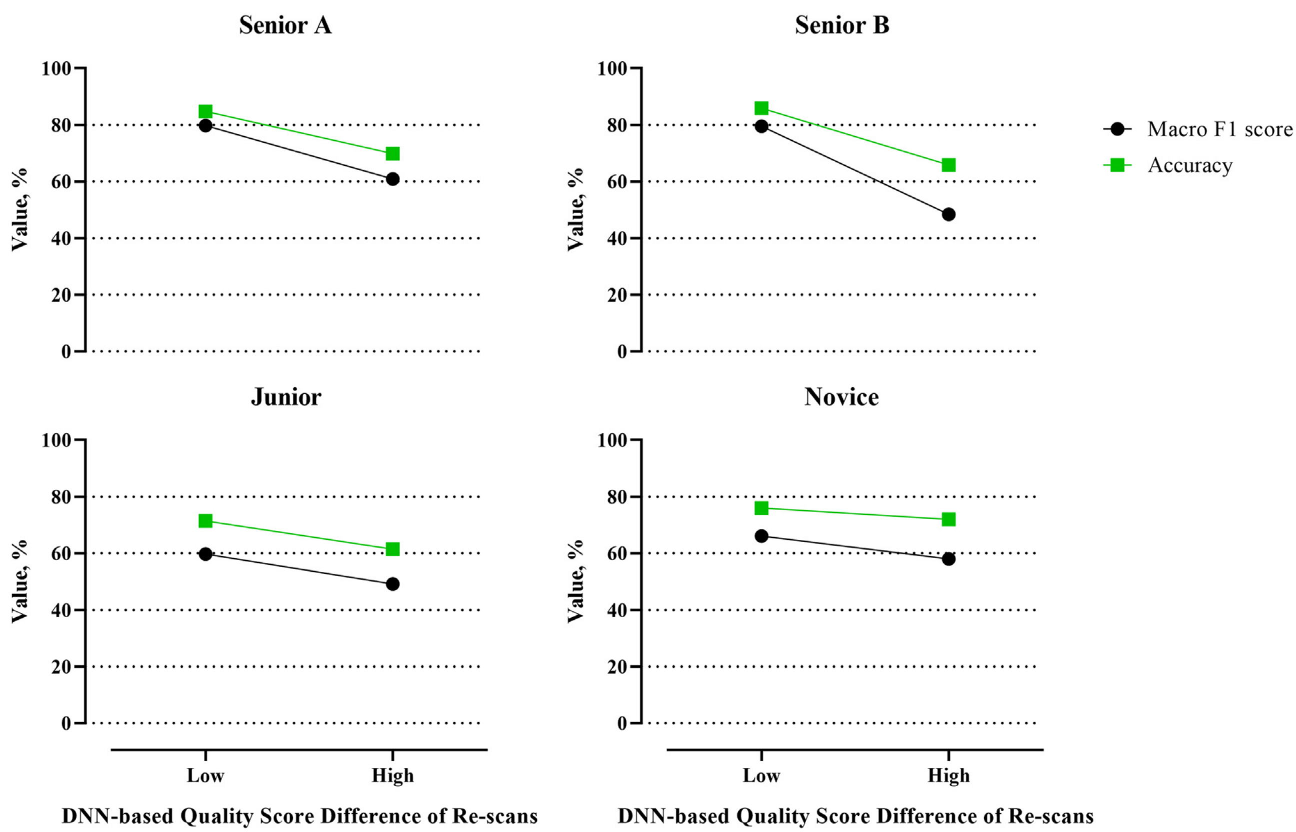

Figure 7. Diagnostic agreement of re-scan sequences in the high- and low-score-difference groups among echocardiographers. The re-scan sequences were equally divided into high-difference (n = 106 pairs) and low-difference (n = 107 pairs) groups. Refer to the color bar for the recall represented by the numbers and colored boxes in the images. A to C: internal validation; D to F: external test. DNN, deep neural network.

implementation of a framework that identifies specific image quality issues and provides specific auto-adjustment prompts is worth considering in future research.

Limitations

Although our study has only achieved limited automation of image quality assessment, we have successfully addressed the most critical factors that influence image quality. However, in future studies, we need to implement quality assessment for gain before injection and other detailed image quality control.

The differences in data distribution between the internal validation and external testing sets in our study are likely due to the limited standardization and adoption of MCE examinations in China. Despite these disparities, the DNN performed well on both data sets, indicating its ability to accurately classify data from new distributions. This suggests that the DNN can capture general patterns and features beyond those seen in the training set. It would be ideal for the DNN to exhibit accurate image quality assessment across a broader range of data distributions and scenarios.

Because the evaluation of flash is limited to only the first frame after it, our sample size for training and testing is relatively small. Moreover, if the DNN classifies flash incorrectly, it can significantly affect the final sequence image quality score. Additionally, the impact of flash qualification on perfusion evaluation is variable, depending on differences in patient circulation rate and body size. For example, even if the image sequence of a small and fast-circulating patient contains excessive flash, it may not cause misdiagnosis of perfusion delay because of the rapid filling of the left ventricle through the circulation. In future research, it will be necessary to use radionuclide perfusion imaging (*e.g.,* single-photon-emission computed tomography) as a gold standard to evaluate the improvement in diagnostic accuracy resulting from image quality control.

It is worth noting that the MCE images used in this study were obtained with the Philips 7c. Therefore, the generalizability of the proposed DNN to other brands of ultrasound machines may be limited, and further validation is necessary.

Conclusion

This study proposed a ResNet18-based framework for real-time automatic MCE image quality assessment and scoring of LVO, shadowing and flash. It was found that the DNN-based image quality scores were associated with perfusion analysis among echocardiographers. We believe that using this DNN for image quality control can enhance the feasibility of perfusion assessment and reduce diagnostic variability among echocardiographers.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 82271999), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021A1515012232) and the National Natural Science Foundation of China launching fund of the Guangdong Provincial People's Hospital (Grant No. 8207070477).

Conflict of interest

The authors declare no competing interests.

Data availability statement

The data reported and codes of this study are available on reasonable request.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2023.07.002.

References

[1] Porter TR, Mulvagh SL, Abdelmoneim SS, Becher H, Belcik JT, Bierig M, et al. Clinical applications of ultrasonic enhancing agents in echocardiography: 2018 American Society of Echocardiography Guidelines Update. J Am Soc Echocardiogr 2018;31:241–74.
[2] Feldman MK, Katyal S, Blackwood MS. US artifacts. Radiographics 2009;29: 1179–89.
[3] Papolos A, Narula J, Bavishi C, Chaudhry FA, Sengupta PP. U.S. hospital use of echocardiography: insights from the nationwide inpatient sample. J Am Coll Cardiol 2016;67:502–11.
[4] Senior R, Becher H, Monaghan M, Agati L, Zamorano J, Vanoverschelde JL, et al. Clinical practice of contrast echocardiography: recommendation by the European Association of Cardiovascular Imaging (EACVI) 2017. Eur Heart J Cardiovasc Imaging 2017;18:1205af.
[5] Porter TR, Abdelmoneim S, Belcik JT, McCulloch ML, Mulvagh SL, Olson JJ, et al. Guidelines for the cardiac sonographer in the performance of contrast echocardiography: a focused update from the American Society of Echocardiography. J Am Soc Echocardiogr 2014;27:797–810.
[6] Cosyns B, Helfen A, Leong-Poi H, Senior R. How to perform an ultrasound contrast myocardial perfusion examination? Eur Heart J Cardiovasc Imaging 2022;23: 727–9.
[7] Deo RC. Machine learning in medicine. Circulation 2015;132:1920–30.
[8] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[9] Xie F, Qian L, Goldsweig A, Xu D, Porter TR. Event-free survival following successful percutaneous intervention in acute myocardial infarction depends on microvascular perfusion. Circ Cardiovasc Imaging 2020;13:e010091.
[10] Aggarwal S, Xie F, High R, Pavlides G, Porter TR. Prevalence and predictive value of microvascular flow abnormalities after successful contemporary percutaneous coronary intervention in acute ST-segment elevation myocardial infarction. J Am Soc Echocardiogr 2018;31:674–82.
[11] Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. IEEE Trans Med Imaging 2017;36: 1221–30.
[12] Abdi AH, Luong C, Tsang T, Jue J, Gin K, Yeung D, et al. Quality assessment of echocardiographic cine using recurrent neural networks: feasibility on five standard view planes. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, editors. Medical image computing and computer assisted intervention—MICCAI 2017. Lecture notes in computer science, 10435. Cham: Springer; 2017.
[13] Luong C, Liao Z, Abdi A, Girgis H, Rohling R, Gin K, et al. Automated estimation of echocardiogram image quality in hospitalized patients. Int J Cardiovasc Imaging 2021;37:229–39.
[14] Dong J, Liu S, Liao Y, Wen H, Lei B, Li S, et al. A generic quality control framework for fetal ultrasound cardiac four-chamber planes. IEEE J Biomed Health Inform 2020;24:931–42.
[15] Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. JAMA Cardiol 2021;6:624–32.