Intelligent Medicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Intelligent Medicine

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

Research Article

A combined system with convolutional neural networks and transformers for automated quantification of left ventricular ejection fraction from 2D echocardiographic images

Mingming Lin[1,#], Liwei Zhang[2,3,#], Zhibin Wang[1], Hengyu Liu[1], Keqiang Wang[4], Guozhang Tang[1], Wenkai Wang[4], Pin Sun[1,*]

[1] *Department of Echocardiography, The Affiliated Hospital of Qingdao University, Qingdao, Shandong 266000, China*
[2] *Molecular Ageing, Helmholtz Pioneer Campus, Helmholtz Center Munich, Neuherberg, Germany*
[3] *Medizinische Klinik und Poliklinik IV, Ludwig-Maximilians-Universität München, Munich, Germany*
[4] *Qingdao Hisense Medical Equipment Co., Ltd., Qingdao, Shandong 266000, China*

ARTICLE INFO

Keywords:
Lightweight model
Transformer
Left ventricular ejection fraction
Device integration
Echocardiography

ABSTRACT

Background Accurate measurement of left ventricular ejection fraction (LVEF) is crucial in diagnosing and managing cardiac conditions. Deep learning (DL) models offer potential to improve the consistency and efficiency of these measurements, reducing reliance on operator expertise.

Objective The aim of this study was to develop an innovative software-hardware combined device, featuring a novel DL algorithm for the automated quantification of LVEF from 2D echocardiographic images.

Methods A dataset of 2,113 patients admitted to the Affiliated Hospital of Qingdao University between January and June 2023 was assembled and split into training and test groups. Another 500 patients from another campus were prospectively collected as external validation group. The age, sex, reason for echocardiography and the type of patients were collected. Following standardized protocol training by senior echocardiographers using domestic ultrasound equipment, apical four-chamber view images were labeled manually and utilized for training our deep learning framework. This system combined convolutional neural networks (CNN) with transformers for enhanced image recognition and analysis. Combined with the model that was named QHAutoEF, a 'one-touch' software module was developed and integrated into the echocardiography hardware, providing intuitive, real-time visualization of LVEF measurements. The device's performance was evaluated with metrics such as the Dice coefficient and Jaccard index, along with computational efficiency indicators. The dice index, intersection over union, size, floating point operations per second and calculation time were used to compare the performance of our model with alternative deep learning architectures. Bland-Altman analysis and the receiver operating characteristic (ROC) curve were used for validation of the accuracy of the model. The scatter plot was used to evaluate the consistency of the manual and automated results among subgroups.

Results Patients from external validation group were older than those from training group ((60±14) years *vs.* (55±16) years, respectively, $P < 0.001$). The gender distribution among three groups were showed no statistical difference (43 % *vs.* 42 % *vs.* 50 %, respectively, $P = 0.095$). Significant differences were showed among patients with different type (all $P < 0.001$) and reason for echocardiography (all $P < 0.001$ except for other reasons). QHAutoEF achieved a high Dice index (0.942 at end-diastole, 0.917 at end-systole) with a notably compact model size (10.2 MB) and low computational cost (93.86 G floating point operations (FLOPs)). It exhibited high consistency with expert manual measurements (intraclass correlation coefficient (ICC) =0.90 (0.89, 0.92), $P < 0.001$) and excellent capability to differentiate patients with LVEF ≥60 % from those with reduced function, yielding an area under the operation curve (AUC) of 0.92 (0.90–0.95). Subgroup analysis showed a good correlation between QHAutoEF results and manual results from experienced experts among patients of different types ($R = 0.93$, 0.73, 0.92, respectively, $P < 0.001$) and ages ($R = 0.92$, 0.94, 0.89, 0.91, 0.81, respectively, $P < 0.001$).

Conclusions Our software-hardware device offers an improved solution for the automated measurement of LVEF, demonstrating not only high accuracy and consistency with manual expert measurements but also practical adaptability for clinical settings. This device might potentially support clinicians and augment clinical decision.

* Corresponding author: Pin Sun, Department of Echocardiography, The Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Shinan District, Qingdao, Shandong 266000, China.
E-mail address: sunpin@qduhospital.cn (Pin Sun).
# These authors contributed equally to this article.

https://doi.org/10.1016/j.imed.2024.10.001
Received 29 June 2024; Received in revised form 15 July 2024; Accepted 17 October 2024
Available online xxx

Please cite this article as: M. Lin, L. Zhang, Z. Wang, et al., QHAutoEF: Development of a combined algorithm with convolutional neural networks and transformers for automated measurement of LVEF from 2D echocardiographic images, Intelligent Medicine, https://doi.org/10.1016/j.imed.2024.10.001

1. Introduction

Quantitative assessment of cardiac systolic function is pivotal for the diagnosis of diseases, risk stratification, and assessment of treatment response [1]. Measurement of left ventricular ejection fraction (LVEF) using echocardiography is generally considered the first-line method for clinical evaluation of left ventricular systolic function [2,3]. In recent years, the advent of artificial intelligence (AI) in the medical image analysis has opened up new avenues and the application of AI in echocardiography has the potential to improve measuring efficiency as well as image interpretation.

Compared with machine learning (ML), deep learning (DL) could study very complex functions through composing several nonlinear modules [4]. Among numerous deep learning models, convolutional neural network (CNN) was the most widely used in medical image analysis and several studies have demonstrated its capability in streamlining workflows, automating chamber segmentation, feature recognition of medical images, and quantifying echocardiographic parameters [5,6]. CNN could learn complex features of echocardiograms through hierarchical structure of convolutional layers. The excellent local feature capture capability of CNNs allows them to interpret precise features of specific diseases from echocardiography images [7–12]. Transformer, an innovative deep-learning architecture, was widely used for natural language processing (NLP) tasks. For medical application, transformer can potentially enhance image analysis tasks by capturing long-range dependencies between image patches in an integrated fashion [13,14]. Previous proposed transformer network allows improved feature extraction, utilizing the inter-frame correlations in echocardiography images [15,16]. CNN-transformer hybrid models were used for multi-task learning, improved image denoising and decreasing overall computational cost in medical fields, mainly in computed tomography and magnetic resonance imaging fields [17–19]. Hybridizing CNNs and transformers is therefore a promising direction for echocardiography tasks. In addition, previous studies focused mainly on model building rather than clinical device application [20,21].

We hypothesized that algorithms hybridizing CNN and transformer can readily evaluate LVEF with high accuracy from 2D echocardiography images. We proposed that such a light-weight software-hardware combined device can be applied in the routine clinical workflows, allowing rapid and accurate assessment of LVEF. In this study, we aimed to establish an echocardiography database from Chinese cohort by domestic equipment and further develop an automatic and accurate LVEF measurement protocol that combines CNN and transformer.

2. Methods

2.1. Preliminary construction of algorithm framework

2.1.1. Algorithm framework

The overall framework of the automated left ventricle segmentation algorithm is illustrated in Figure 1A. It comprises an Encoder backbone and two head networks. Given an echocardiography image as input, Encoder backbone first extracts multilevel features, then uses feature pyramids to segment images from coarse to fine through two head networks. In the fine-aligned branch, a spatial channel attention-based encoder (CSE) module extracts information on mask-alignment from fine feature maps and then feeds it back to coarse-alignment branch. The enhanced coarse-aligned mask was processed by a feature-positioning (FP) module.

2.1.2. Channel-spatial encoder module

The transformer architecture was used to better segment cardiac chambers using structured information from ultrasound images. Given the input fine-aligned feature map, the CSE module first deforms the feature map F to obtain query Q, key K and value V. Subsequently, the module establishes long-term dependencies on the channel and space of feature map from a global perspective, enhancing the semantic information of fine-aligned branch feature map. Finally, an initial profile of the cardiac chamber was obtained by applying convolution layers (Figure 1B).

2.1.3. Feature-positioning module

Due to the unique imaging principle of ultrasound, some degree of acoustic noise will be present in the image and some noise will appear in the observed area, which will affect the scanning results. When performing an ultrasound examination, the physician neglects the noise of the ultrasound image based on clinical experience in order to outline the chamber of the heart and make the final diagnosis. For this reason, we try to mimic this behavior with different network structures, excluding false positives or false negative interferences that predict regional heterogeneity. Therefore, the FP module was designed to fuse features at different levels, as shown in Figure 1C. For feature maps at different scales from two head networks, the FP module is guided by the fine-aligned branch predicted cardiac chamber contours figure, using element-by-element subtraction operations to suppress noise interference and element-by-element summation to enhance target foreground information, respectively. A more precise prediction map of the chamber of the heart was obtained by applying a convolution layer on the refined fusion feature map.

2.1.4. Horizontal comparison of algorithms

In order to investigate whether this model was superior compared to prior models, we assembled three neural network models and selected four widely used segmentation networks for lateral comparison in public datasets. (1) U-Net structure: a classical CNN model that greatly improves the accuracy of pixel level semantic segmentation through the jumping connection between encoder and decoder; (2) U-Net^{++}: an advanced structure that foster better inter-layer information transfer compared with U-Net, which improves the sensitivity of the model to different sizes of receptive fields; (3) Attention Unet: an encoder-decoder structure with attention gating module for medical image segmentation, increases the weight of key parts in the region by increasing the attention mechanism in the decoder, making the model converge faster while increasing the sampling accuracy; (4) ACNN: a novel training strategy for cardiac image segmentation, constraining the model through the prior knowledge of the anatomical structure to improve the segmentation accuracy during training. The CAMUS heart dataset is used for horizontal comparison, and finally the self-built and top-performing model on the CAMUS dataset will be selected as the model adopted in this protocol.

2.2. Database establishment

A dataset with patients who underwent echocardiography at three campuses (the Shinan campus, the Pingdu campus and the west coast campus) of the Affiliated Hospital of Qingdao University from January to June 2023 was established prospectively for model training and internal test groups. For the external validation of the model, 500 echocardiography videos corresponding to 500 patients were prospectively obtained from July to August 2023, mirroring real clinical environments at the Laoshan campus of the Affiliated Hospital of Qingdao University. The exclusion criteria for the study population included: (1) age <18 years; (2) patients with very poor image quality so that the endocardium could not be completely identified; (3) patients with thoracic deformity; (4) images with foreshortening. The reason for echocardiography (physical examination, rehabilitation and health care, preoperative evaluation, cardiovascular disease evaluation, prenatal examination, non-cardiovascular disorders, radiotherapy and other reasons) and the type of patients (outpatient, inpatient or emergency) were collected.

All study participants provided informed consent, and the study design was performed in line with the Declaration of Helsinki and approved by the ethics committee of the Affiliated hospital to Qingdao University (Approval No. QYFYEC2024–34). All methods in our study

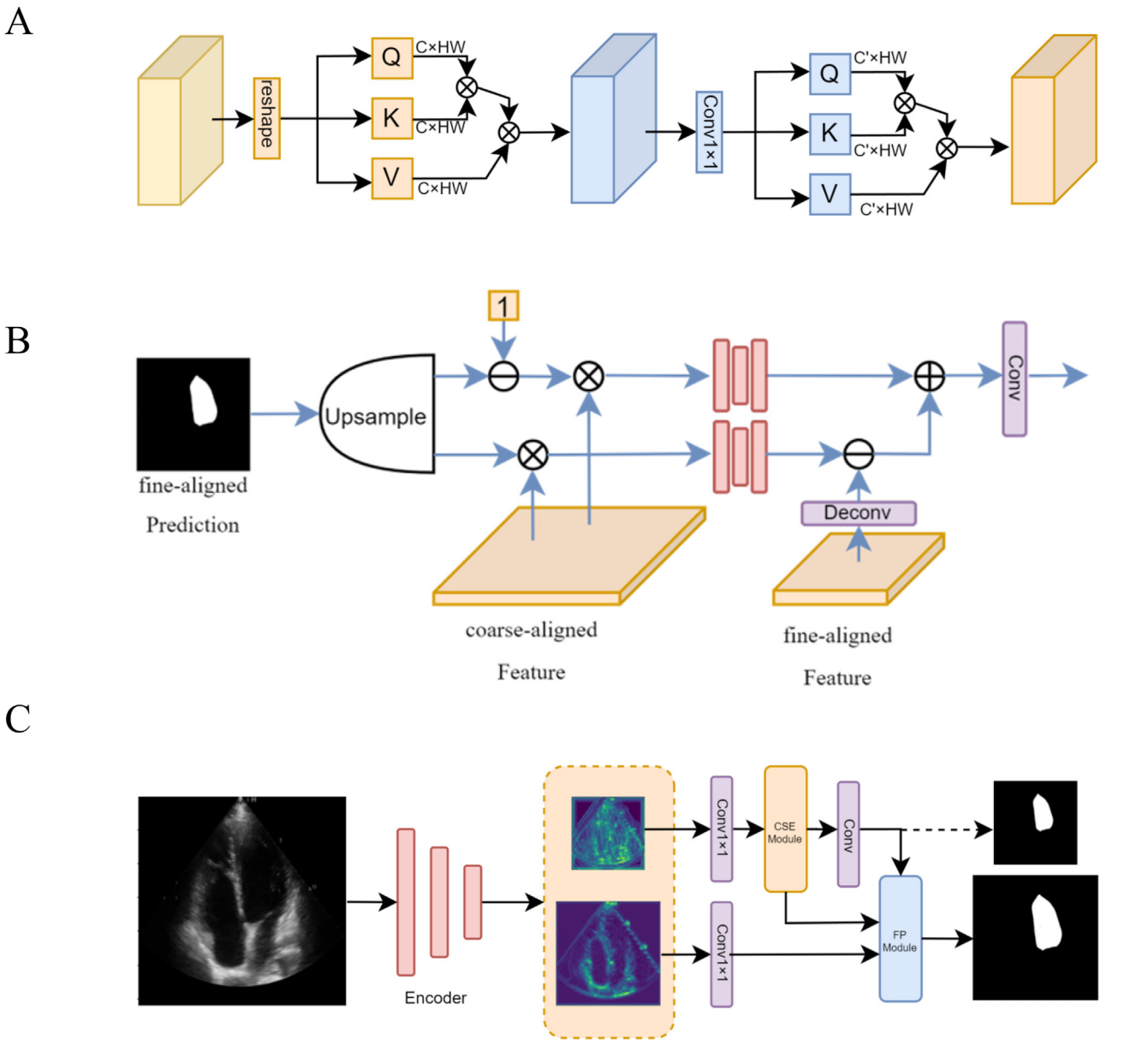

Fig. 1. (A) Overall block diagram of automated chamber segmentation algorithm; (B) The Channel-Spatial Encoder Module using transformer; (C) Feature-Positioning Module.

were carried out in accordance with relevant guidelines and regulations in the Ethics approval.

The original data are not publicly accessible as the model is intended for commercial release. The data that supports the findings of this study is available from the corresponding author upon prior request.

Before database establishment, all of researchers received standardized training, including image acquisition process, image storage process, data interpretation procedure, and image labeling protocol according to established guidelines for echocardiography [22,23]. Patients were all performed echocardiography twice by the same senior echocardiographer. LVEF values measured using a commercial echocardiography machine (Philips Epiq 7C with S5–1 probe) were saved as the gold standard. And then all images were collected using Hisense echocardiographic system (HD80 with P4–1EL probe) for training and test of the algorithm. Four dynamic videos were collected for each patient, including parasternal left ventricular long-axis view (PLAX), apical four-chamber view (A4C), apical two-chamber view (A2C), and apical three-chamber view (A3C), each of which contained 6–10 cardiac cycles. Each patient was anonymized and assigned to a set of echocardiographic image datasets.

2.3. Algorithm training and testing

From the database, the A4C videos were extracted to get the training dataset for model training and validation. The dynamic videos were converted to static images as input dataset using Hisense ultrasound dedicated software. The end-diastolic and end-systolic frames of each cardiac cycle were identified by a senior echocardiographer and the endocardial borders were manually traced with LabelMe, a tool created by the Massachusetts Institute of Technology.

The optimal model, selected out of the seven test models, was used for training. The PyTorch framework was used for code implementation of model training and validation. The hardware environment was Intel(R) Core (TM) i7–10700F CPU @ 2.90 GHz and NVIDIA GeForce RTX 3090 GPU (24 G). In the training phase, we performed an automated pre-processing workflow to remove identifying information and human

labeling. All subsequent videos were cropped and masked to remove any text, ECG and respirator information. Normalized 512 × 512 pixel single channel images were obtained by bilinear interpolation using OpenCV. Data expansion was performed by PyTorch library for up-and-down flip, left-right flip, and random rotation.

The optimizer was Adam method with an initial learning rate of 0.0015 and a weight decay of 0.3. The batch size was 32 and the training was performed for 300 cycles in which the first 15 cycles were taken as warmup stages.

2.4. Algorithm external validation

500 videos from the external validation group were saved for the performance validation of the model. The Bland-Altman analysis and the receiver operating characteristic (ROC) curve were used for validation of the accuracy of the model. The scatter plot was used to evaluate the consistency of the manual and automated results among subgroups.

2.5. Software development and software-hardware integration

This novel model was named QHAutoEF in which "Q" means the Affiliated Hospital of Qingdao University and "H" means Hisense Medicine. The PyTorch library facilitated the storage of QHAutoEF as a model file (.pt format) containing the network structure and all weight parameters. The .pt file was serialized and converted into the high-performance DL support engine, TensorRT, to enable seamless integration. During serialization, we employed the half-precision floating-point format for data representation to reduce model size. In addition, we developed inference code as dynamic link libraries, which were then integrated and compiled with the standard echocardiography control software. The frame information was bonded to the knob for visualization output.

In the real clinical settings, echocardiographer activated QHAutoEF through the specialized button and the one-touch visualization functionality proceeded through the following sequence: (a) loading of the QHAutoEF model when the echocardiography machine was powered on; (b) de-constructing echocardiography videos into individual frames, each of which was transferred to the graphics processing unit (GPU) from the central processing unit (CPU). This was followed by model inference, after which the results were sent back to the CPU; (c) combining the inference outcomes with the original echocardiography images for display on the monitor; (d) manually adjusting the visualization, frame by frame using the control knob; (e) generating a volume-time curve with a single button press; (f) clearing all memory and cache, leaving the model ready for the next activation.

2.6. Data protection

All the collected data was converted to DICOM format. During the pre-processing stage, only the original image data was extracted from the DICOM dataset and saved in PNG format, without including any patient-identifying information. In addition, all data handling and model training processes took place in a secure offline environment to ensure that no data was uploaded to external servers.

2.7. Statistical analysis

To assess the distribution of the measured data, a normality test was applied. Data conforming to normal distribution was represented as mean ± standard deviation (SD), whereas non-normally distributed data was described using the median and interquartile range. Categorical variables were expressed as frequency and percentage. According to the sample distribution, variance, and experimental setting, we used parametric independent samples *t*-test or Pearson's chi-quare test to test against differences between groups. Several metrics were calculated to evaluate the performance of the QHAutoEF tool: (a) Dice index, model size; (b) floating point operations (FLOPs); (c) calculation time; (d) mean squared error (MSE); (e) root mean squared Error (RMSE); (f) mean absolute error (MAE); (g) mean absolute percentage Error (MAPE); (h) relative absolute error (RAE); (i) relative squared error (RSE); (j) adjusted coefficient of determination (R^2_{adj}); (k) Intersection over Union (IoU). The consistency between QHAutoEF measurements and senior echocardiographer measurements was assessed using Bland-Altman analysis and intraclass correlation coefficient (ICC). ROC curves and area under the curve (AUC) were employed to demonstrate QHAutoEF's capacity to accurately categorize patients with varying levels of LVEF. Statistical significance was set at a P value <0.05. Statistical analysis used IBM SPSS statistics version 26 (IBM Corporation, Armonk, NY, USA), Graphpad Prism 9, and draw.io online platforms.

3. Results

3.1. Baseline data of the study population

In this study, 2,613 echocardiograms, each corresponding to a unique patient, were prospectively incorporated into the analysis. Out of these, 2,113 echocardiograms (1,913 from individual patients for model training and 200 for internal testing) were allocated to the development of the predictive model. The remaining 500 echocardiograms were designated for external validation of the model. The baseline characteristics of the study population are shown in Table 1. Compared to training and test group, patients from external validation group were older ((60±14) years *vs.* (55±16) years *vs* (59±14) years, respectively in validation, training and test group, $P < 0.001$). No statistical difference was showed in the gender distribution among three groups (43 % *vs* 42 % *vs* 50 %, respectively, $P = 0.095$). Significant differences were showed among patients with different type (all $P < 0.001$). The percentage of patients for the reason other than common 7 reasons for echocardiography showed no statistical difference among groups (2 % *vs.* 1 % *vs.* 2 %, respectively, $P = 0.367$). The distribution of patients with common reasons for echocardiography showed significant difference among training, test and validation groups (all $P < 0.001$).

3.2. Performance of QHAutoEF

QHAutoEF exhibits superior characteristics including a smaller size (10.2 MB), lower floating-point operations (93.86 G), fast calculation (22 ms) and higher Dice index and intersection over union ($Dice_{ED}$=0.942, $Dice_{ES}$=0.917, IoU_{ED}=0.89, IoU_{ES}=0.85) compared with other models for automatic measurement of LVEF (Table 2).

3.3. The accuracy validation of QHAutoEF

QHAutoEF demonstrated strong performance in the external validation cohort, achieving $Dice_{ED}$=0.93, $Dice_{ES}$=0.91, MAE=3.67 %, MSE=20.31 %, RMSE=4.51 %, MAPE=6.72 %, RSE=0.21, RAE=0.62, R^2_{adj}=0.82, IoU_{ED}=0.87 and IoU_{ES}=0.84. There was a high degree of consistency between the automatic measurements made by QHAutoEF and the manual measurements conducted by experts, as indicated by ICC of 0.90 (0.89, 0.92) and a statistically significant p-value ($P < 0.001$). Compared with manual measurement, QHAutoEF measurement was marginally lower with a difference of 1.119 % (95 % CI: −7.447, 9.686) (Figure 2A). For distinguishing patients with LVEF ≥ 60 % from those with LVEF < 60 %, the AUC for QHAutoEF was 0.92 (0.90–0.95), including sensitivity of 0.88 and a specificity of 0.80 (Figure 2B)

In Figure 2C, we present illustrative results showcasing the performance QHAutoEF. The visual representation clearly demonstrates that our proposed QHAutoEF showed high consistency with ground truth in terms of visual quality, especially when trained using domestic Chinese cohort.

Subgroup analysis showed a good correlation between QHAutoEF results and manual results from experienced experts in patients from

Table 1 Patient characteristics in different dataset

Statistics	Total	Training	Test	Validation	*P* -values
Number of Patients (*n*)	2,613	1,913	200	500	
Age (years (±SD))	56 ± 16	55 ± 16	59 ± 14	60 ± 14[#]	<0.001
Male (*n* (%))	1,163 (44)	828 (43)	85 (42)	250 (50)	0.095
Patient types (*n* (%))					<0.001
Emergency	514 (20)	418 (22)	17 (9)	79 (16)	<0.001
Outpatient	994 (38)	844 (44)	43 (22)	107 (21)	<0.001
Inpatient	1,105 (42)	651 (34)	140 (70)	314 (63)	<0.001
Reason for presentation					<0.001
RHC (*n* (%))	180 (7)	161 (8)	13 (6)	6 (1)	<0.001
Prenatal examination (*n* (%))	80 (3)	23 (1)	2 (1)	55 (11)	<0.001
Preoperative evaluation[†] (*n* (%))	975 (37)	654 (34)	141 (71)	180 (36)	<0.001
Physical examination (*n* (%))	503 (19)	371 (19)	13 (7)	119 (24)	<0.001
Radiotherapy[‡] (*n* (%))	197 (8)	175 (9)	8 (4)	14 (3)	<0.001
Cardiovascular disease (*n* (%))	211 (8)	180 (9)	5 (3)	26 (5)	<0.001
Non-cardiovascular disorders (*n* (%))	421 (16)	313 (16)	17 (9)	91 (18)	0.006
Other (*n* (%))	46 (2)	36 (2)	1 (1)	9 (2)	0.367

RHC: rehabilitation and health care.
[†] pre-operative evaluation for non-cardiovascular surgery.
[‡] evaluation of cardiac structure and function after radiotherapy.

Table 2 QHAutoEF performance compared to alternative deep learning architectures in assessing cardiac function

Methods	Dice index		IoU		Size (MB)	FLOPs (G)	Calculation time (ms)
	ED	ES	ED	ES			
U-Net [34]	0.94	0.91	0.88	0.84	2	16.59	16
U-Net [++] [35]	0.92	0.90	0.86	0.82	9.5	86.55	40
Attention Unet [36]	0.92	0.91	0.86	0.83	30.2	266.54	57
ACNN [37]	0.94	0.91	0.88	0.84	16.8	158.45	42
Model 1	0.94	0.92	0.88	0.84	17.2	137.22	33
Model 2 (QHAutoEF)	0.94	0.92	0.89	0.85	10.2	93.86	22
Model 3	0.91	0.88	0.83	0.78	6.8	45.30	17

ED: end-diastole; ES: end-systole; FLOPs: floating point operations per second; IoU: Intersection over Union; MB: megabyte; G: 10^9.

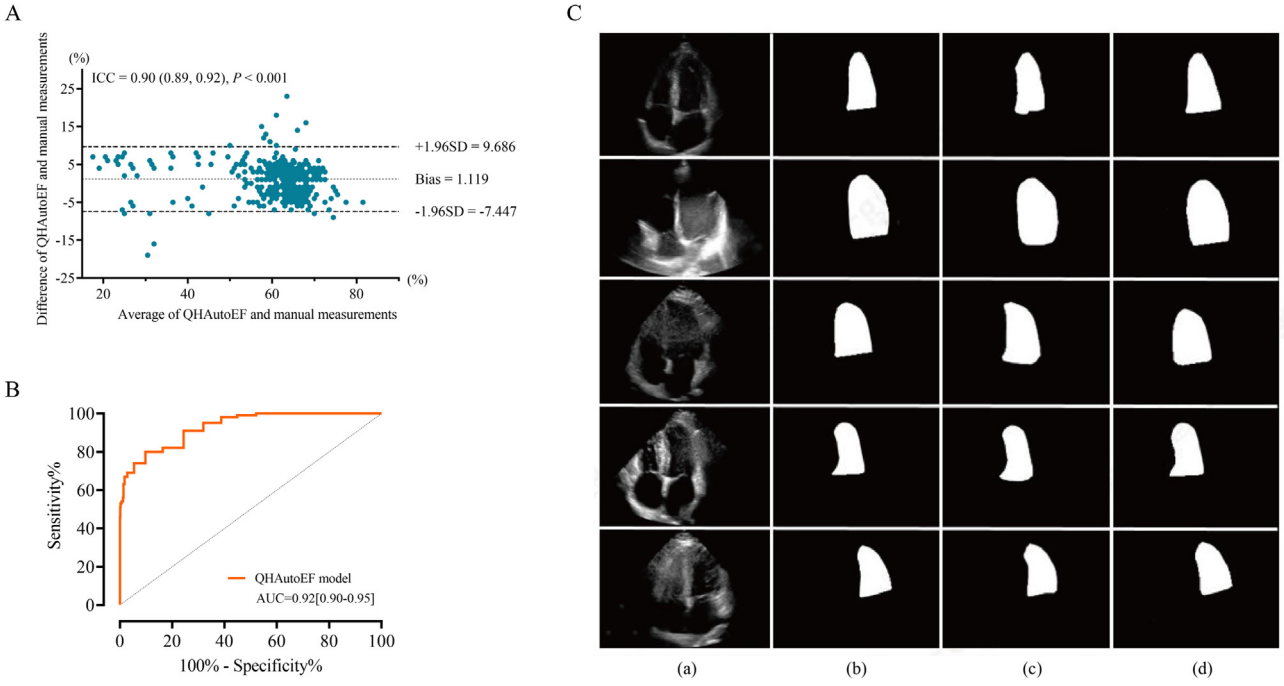

Fig. 2. QHAutoEF performance included consistency analysis (A), ROC analysis (B), visualization results (C), (a) original echocardiography images; (b) ground truth; (c) training on CAMUS datasets and validation on Chinese cohort; (d) training on CAMUS plus Chinese datasets and validation on Chinese cohort. ICC: intra-class correlation coefficient; AUC: area under the curve.

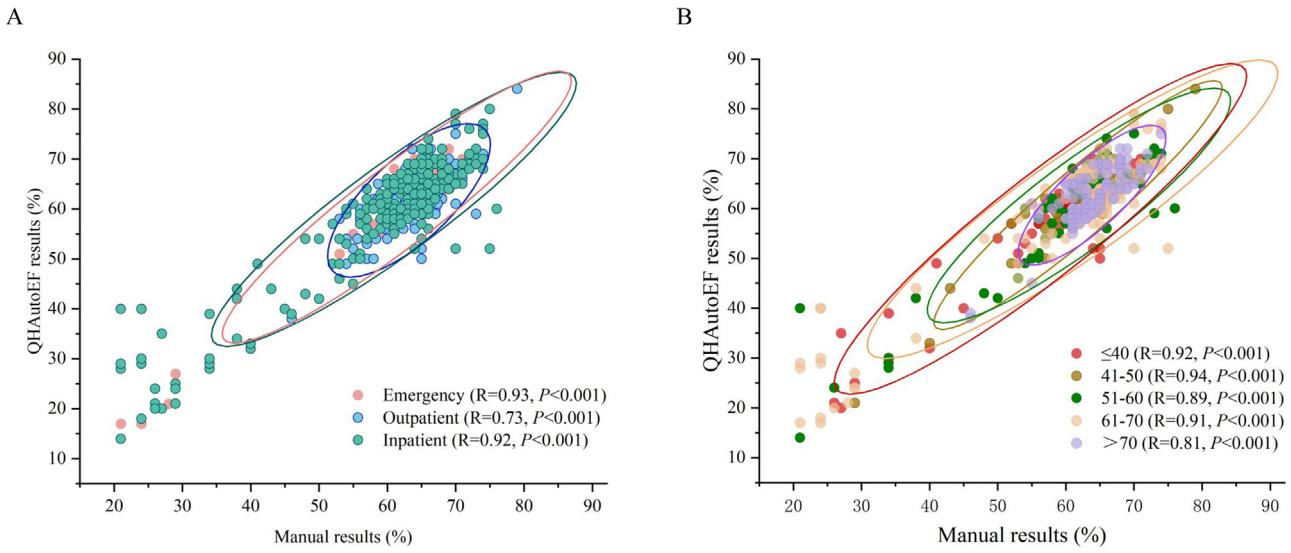

Fig. 3. Subgroup analysis for the association between QHAutoEF results and experts' manual results. (A) Subgroup analysis according to different origins; (B) Subgroup analysis according to different age.

the emergency department, outpatient clinic, and inpatient settings (R = 0.93, 0.73, 0.92, respectively, P < 0.001. Figure 3A). When the validation cohorts were divided into subgroups by age, high consistency was obtained among patients with different levels of age (R = 0.92, 0.94, 0.89, 0.91, 0.81, respectively, P < 0.001. Figure 3B).

3.4. Visualization output by the one-touch module

The one-touch module facilitates three key functions: (1) delicate adjustment: precise frame-by-frame adjustment is possible by rotating the knob, allowing for accurate measurement of LVEF. (2) cardiac cycle switching: the module automatically presents the varying LVEF values of different cardiac cycles on the screen with a simple turn of the knob. (3) Volume-Time Curve: a comprehensive volume-time curve for a complete cardiac cycle is displayed through one-touch operation, enabling the observation of left ventricular volume changes throughout the cycle.

Two cases are illustrated in Figure 4. In case 1 (Figure 4A), the LVEF measured by the expert was 65 %. QHAutoEF provided a slightly higher result of 69 %. However, after a delicate adjustment by one frame, the LVEF was re-adjusted to 66 %. In case 2 (Figure 4B), the LVEF measured by the expert was 65 %. However, the views captured in one of the three cycles were not standard (the first cycle). Consequently, by uti-

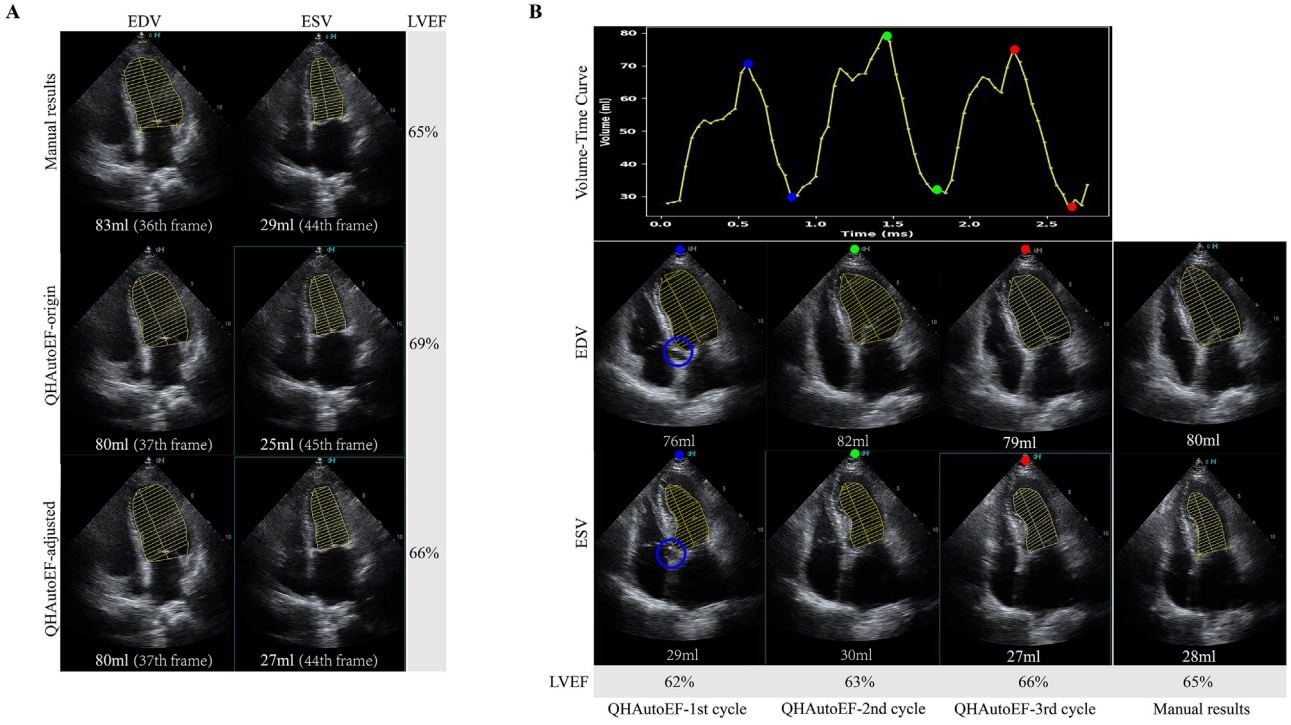

Fig. 4. Illustration of results obtained manually and via QHAutoEF output in 2 example cases. Fig. 4A showed more accurate results from delicate adjustments and Fig. 4B showed cardiac cycle switching to avoid inaccurate results from non-standard views. The blue circle in Fig. 4B marked aortic valve that should not appear in the standard apical four chamber view.

lizing the cardiac cycle switching knob, the LVEF from alternative cycles were promptly displayed at 63 % (the second cycle) and 66 % (the third cycle). This switching functionality was enabled by paired EDV and ESV values obtained from the volume-time curve. The operation penal with special buttons for QHAutoEF was showed in Supplementary Figure 1 and the software interface was showed in Supplementary Figure 2.

4. Discussion

In this study, we developed a Chinese software-hardware device with a light-weight hybrid CNN-transformer model from echocardiography images obtained from a Chinese cohort. This model, named QHAutoEF, is capable of performing quick and automatic measurements of LVEF with good accuracy.

Artificial intelligence, especially DL used in medical field, has greatly improved the efficiency of diagnosis and treatment in a variety of medical scenarios [8,20,24,25]. CNNs were the primary methods for interpretation of echocardiograms. Previous studies have demonstrated feasibility of CNNs for standard views classification [6], segmentation of cardiac chambers [5], assessment of systolic function [21], specific diseases detection [8,10,11,24], and prediction of cardiovascular risk factors or outcomes [9,26]. Madani's group established a CNN model based on 223,787 echocardiographic static images from 267 patients, which showed that the model accurately identified 15 standard views for echocardiography [6]. Ghorbani developed a CNN model based on 2,600,000 images from 2,850 patients and this model was used to accurately identify cardiac pacemaker leads and estimate left ventricular end-systolic and end-diastolic volumes and further calculate LVEF [5]. Notably, these models often have a large number of parameters in order to achieve high predictive accuracy. This complexity however, can lead to practical challenges in clinical deployment due to computational resources required, longer inference times, and difficulties in integrating it with the existing healthcare systems.

The Transformer architecture represents an innovative approach that can be employed in medical imaging tasks, which relies primarily on self-attention mechanisms to discern dependencies and associations within sequences of data [27]. In the analysis of biomedical images, the Transformer conceptualizes each image patch as an individual element in a sequence, processing the image as a contiguous array of such patches. This methodology facilitates the capture of long-range dependencies among patches and effectively models spatial relationships within the image. By leveraging self-attention, it markedly enhances classification performance across a wide spectrum of image analysis tasks [13,14,28,29].

Given the constraints of CNNs in capturing global information and the shortcomings of transformer models in extracting local details, hybrid CNN-transformer models have been developed to improve performance [30,31]. These models commonly employ either a serial or parallel integration strategy to merge CNN and transformer architectures, yielding superior results compared to the individual performances of CNNs or transformers. Two-dimensional features of input images were subsequently refined and encoded with positional information before being further learned by the transformer, resulting in more accurate classification in Carion's study [32]. The ViT-FRCNN architecture, presented by Beal, integrates a faster R-CNN in series following the vision transformer (ViT), showcasing the transformer's capability to preserve ample spatial information for target detection [33]. Unlike the above methods, the QHAutoEF framework presented in this study, employs a CNN as its backbone, while the transformer serves as one aspect of a dual-branch system. This transformer interlaces the temporal and spatial dimensions of the image features to establish branch-specific attributes. These are subsequently amalgamated with the local features discerned by the CNN via a feature-positioning module. The process culminates in the formation of comprehensive segmentation results. Notably, this fusion of CNN and transformer architectures demonstrates potential to yield highly fitting performances, producing commendable outcomes even in the analysis of low-quality images.

Previous studies have primarily concentrated on the development of models with exceptional performance, neglecting the creation of practical devices that could be broadly utilized in clinical settings. The QHAutoEF architecture is characterized by its automatic nature, lightweight design, compactness, and minimal computational resource requirements. Consequently, it is suitable for integration with echocardiography equipment possessing modest hardware specifications as it is adaptable for use across diverse medical institutions with varying equipment capabilities. The QHAutoEF is particularly beneficial for remote areas with limited resources, as well as in health screening centers for disease detection. In our device, dynamic link libraries (DLLs) facilitate improved interaction between the hybrid model and hardware. This integration simplifies processes and enables rapid display of results through direct memory level manipulation using DLLs. In addition, DLLs standardize the interface between the model and the echocardiography software in scenarios that require extensive deployment. This standardization ensures that consistent and accurate results are readily obtainable by adhering to the interface protocols, thereby enhancing the repeatability of the deployment process.

The automatic measurement of LVEF using QHAutoEF required only apical four-chamber view, which was helpful for greenhorns to quickly master the accurate measurements. This might greatly increase the efficiency of cardiac function evaluation, which significantly shortens the patient waiting times and thus improves the patient's experience. The automated measurements of LVEF by QHAutoEF demonstrated a strong concordance with expert manual measurements, aligning with the results of previous research. This is in agreement with the findings of Ouyang [21], who utilized a 3D-CNN architecture that incorporated both temporal and spatial parameters for LVEF evaluation. However, the substantial size of the 3D-CNN model necessitates significant GPU acceleration, suggesting that its application is typically confined to advanced medical centers and research facilities.

According to the American Society of Echocardiography (ASE) guidelines, the assessment of cardiac function in patients with arrhythmia necessitates a comprehensive evaluation through multi-cycle averaging, with the mean LVEF representing patient cardiac function [22]. While this protocol is accurate, it is time-consuming. In contrast, QHAutoEF can compute the LVEF for each cardiac cycle, presenting the data through its visualization module. The visualization module enables the echocardiographers to select cardiac cycles easily with a single click. Additionally, in cases with poor image quality, repeated dynamic observation is often essential for precise endocardial delineation, even for experienced sonographers in clinical practice. The frame-by-frame endocardial outline provided by QHAutoEF could significantly enhance the accuracy and efficiency of endocardial identification in such instances. In this study, the QHAutoEF architecture was developed based on a substantial sample from the Chinese population, incorporating patients with various cardiovascular conditions, thus allowing this streamlined model to achieve accurate measurement of LVEF in different levels.

In this study, we collected patients from 3 campuses (Shinan, Pingdu and West coast) for training and test while patients from another campus (Laoshan) for external validation of the model. Actually, the 4 campuses were located in different regions and were far apart, and the functional positioning of each campus was also completely different. Consequently, patients collected from the four campuses could be seen as different groups with significant heterogeneity. And this was verified from our results. For echocardiography in clinical, a measurement with higher objectiveness might be more valuable for future development. Our results showed QHAutoEF owns the potential to reduce subjectivity in LVEF measurement which implied its application prospect in clinical. In addition, knowledge graph can be used for update of QHAutoEF according to different guidelines. A knowledge base including different guideline criteria can serve as an intermediary before final output. Future studies should be able to investigate the use of reinforce-

ment learning to assist sonographers in acquiring images of sufficient quality.

The clinical value of QHAutoEF still needs further validation in different subgroups. In this study, we aim to introduce the independent development process of this localized model, and our results show that QHAutoEF has a high consistency with manual measurement by experts, though with a relatively lightweight size. In addition, the training of QHAutoEF only used images from Hisense devices, but the model could successfully recognized .JPG images derived from commercial devices. QHAutoEF was not integrated into other commercial devices, which was more because of the confidentiality among different vendors. However, this study aims to establish a domestic device that integrated with the deep learning model, the application of QHAutoEF among commercial echocardiography devices might be analyzed in our further study.

In this study, we established a database of a representative Chinese population using domestic equipment and then independently developed a software-hardware device featured hybrid CNN-transformer model. This device offers automated measurements of LVEF with high accuracy, which can support clinicians and augment clinical care.

Conflicts of interest statement

All authors declare no competing interests.

Funding

The authors received no financial support for this research.

Author contributions

All authors contributed equally to critical revisions of the Article and all authors accept final responsibility for the decision to submit for publication. Mingming Lin, Chunyao Liu, Hengyu Liu, Guozhang Tang, Zhibin Wang and Pin Sun established the database, Mingming Lin performed statistical analyses. Mingming Lin wrote the first draft of the Article with input from Keqiang Wang and Wenkai Wang. The study was designed by Pin Sun, Zhibin Wang and Liwei Zhang. Critical revisions of the Article were provided by Liwei Zhang, Zhibin Wang and Pin Sun. All authors had access to all the data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.imed.2024.10.001.

References

[1] Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol 2016;13:368–78. doi:10.1038/nrcardio.2016.25.
[2] Authors/Task Force M, McDonagh TA, Metra M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Eur J Heart Fail 2022;24:4–131. doi:10.1002/ejhf.2333.
[3] Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: a Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2017;136:e137–61. doi:10.1161/CIR.0000000000000509.
[4] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88. doi:10.1016/j.media.2017.07.005.
[5] Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. NPJ Digit Med 2020;3:10. doi:10.1038/s41746-019-0216-8.
[6] Madani A, Arnaout R, Mofrad M, et al. Fast and accurate view classification of echocardiograms using deep learning. NPJ Digit Med 2018;1:6. doi:10.1038/s41746-017-0013-1.
[7] Edwards LA, Feng F, Iqbal M, et al. Machine learning for pediatric echocardiographic mitral regurgitation detection. J Am Soc Echocardiogr 2023;36:96–104 e104. doi:10.1016/j.echo.2022.09.017.
[8] Krishna H, Desai K, Slostad B, et al. Fully automated artificial intelligence assessment of aortic stenosis by echocardiography. J Am Soc Echocardiogr 2023;36:769–77. doi:10.1016/j.echo.2023.03.008.
[9] Lau ES, Di Achille P, Kopparapu K, et al. Deep learning-enabled assessment of left heart structure and function predicts cardiovascular outcomes. J Am Coll Cardiol 2023;82:1936–48. doi:10.1016/j.jacc.2023.09.800.
[10] Sun D, Hu Y, Li Y, et al. Chamber Attention Network (CAN): towards interpretable diagnosis of pulmonary artery hypertension using echocardiography. J Adv Res 2024;63:103–15. doi:10.1016/j.jare.2023.10.013.
[11] Yuan N, Kwan AC, Duffy G, et al. Prediction of coronary artery calcium using deep learning of echocardiograms. J Am Soc Echocardiogr 2023;36:474–81 e473. doi:10.1016/j.echo.2022.12.014.
[12] Wegner FK, Benesch Vidal ML, Niehues P, et al. Accuracy of deep learning echocardiographic view classification in patients with congenital or structural heart disease: importance of specific datasets. J Clin Med 2022;28:690. doi:10.3390/jcm11030690.
[13] Gheflati B, Rivaz H. Vision transformers for classification of breast ultrasound images. Annu Int Conf IEEE Eng Med Biol Soc 2022;2022:480–3. doi:10.1109/EMBC48229.2022.9871809.
[14] Liu Z, Lv Q, Yang Z, et al. Recent progress in transformer-based medical image analysis. Comput Biol Med 2023;164:107268. doi:10.1016/j.compbiomed.2023.107268.
[15] Ahn SS, Ta K, Thorn SL, et al. Co-attention spatial transformer network for unsupervised motion tracking and cardiac strain analysis in 3D echocardiography. Med Image Anal 2023;84:102711. doi:10.1016/j.media.2022.102711.
[16] Ahmadi N, Tsang MY, Gu AN, et al. Transformer-based spatio-temporal analysis for classification of aortic stenosis severity from echocardiography cine series. IEEE Trans Med Imaging 2024;43:366–76. doi:10.1109/TMI.2023.3305384.
[17] Zhou T, Zhang X, Lu H, et al. GMRE-iUnet: isomorphic Unet fusion model for PET and CT lung tumor images. Comput Biol Med 2023;166:107514. doi:10.1016/j.compbiomed.2023.107514.
[18] Yuan J, Zhou F, Guo Z, et al. HCformer: hybrid CNN-transformer for LDCT image denoising. J Digit Imaging 2023;36:2290–305. doi:10.1007/s10278-023-00842-9.
[19] Cheng J, Liu J, Kuang H, et al. A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping. IEEE Trans Med Imaging 2022;41:1520–32. doi:10.1109/TMI.2022.3142321.
[20] Zhang Z, Zhu Y, Liu M, et al. Artificial intelligence-enhanced echocardiography for systolic function assessment. J Clin Med 2022;11:2893. doi:10.3390/jcm11102893.
[21] Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature 2020;580:252–6. doi:10.1038/s41586-020-2145-8.
[22] Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28:1–39. doi:10.1016/j.echo.2014.10.003.
[23] Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 2019;32:1–64. doi:10.1016/j.echo.2018.06.004.
[24] Laumer F, Di Vece D, Cammann VL, et al. Assessment of artificial intelligence in echocardiography diagnostics in differentiating takotsubo syndrome from myocardial infarction. JAMA Cardiol 2022;7:494–503. doi:10.1001/jamacardio.2022.0183.
[25] Upton R, Mumith A, Beqiri A, et al. Automated echocardiographic detection of severe coronary artery disease using artificial Intelligence. JACC Cardiovasc Imaging 2022;15:715–27. doi:10.1016/j.jcmg.2021.10.013.
[26] Tromp J, Bauer D, Claggett BL, et al. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. Nat Commun 2022;13:6776. doi:10.1038/s41467-022-34245-1.
[27] Vaswani A, Shazeer NM, Parmar N, et al. Attention is All you Need. In: Proceedings of 31st Annual Conference on Neural Information Processing Systems. NIPS; 2017.
[28] Vickovic S, Eraslan G, Salmén F, et al. High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods 2019;16:987–90. doi:10.1038/s41592-019-0548-y.
[29] Sagar A. ViTBIS: Vision transformer for biomedical image segmentation. Cham: Springer International Publishing; 2021. p. 34–45.
[30] He Q, Yang Q, Xie M. HCTNet: a hybrid CNN-transformer network for breast ultrasound image segmentation. Comput Biol Med 2023;155:106629. doi:10.1016/j.compbiomed.2023.106629.
[31] Gao G, Xu Z, Li J, et al. CTCNet: a CNN-transformer cooperation network for face image super-resolution. IEEE Trans Image Process 2023;32:1978–91. doi:10.1109/TIP.2023.3261747.
[32] Carion N, Massa F, Synnaeve G. End-to-End object detection with transformers. Cham: Springer International Publishing; 2020. p. 213–29.
[33] Beal J, Kim E, Tzeng E, et al. Toward transformer-based object detection; 2020. doi:1048550/arXiv201209958.
[34] Leclerc S, Smistad E, Pedrosa J, et al. Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography. IEEE Trans Med Imaging 2019;38:2198–210. doi:10.1109/TMI.2019.2900516.
[35] Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. UNet + +: a Nested U-net architecture for medical image segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support 2018;11045:3–11. doi:10.1007/978-3-030-00889-5_1.
[36] OO JS, Folgoc LL, Lee M, et al. Attention U-Net: learning where to look for the pancreas; 2018. doi:1048550/arXiv180403999.
[37] Oktay O, Ferrante E, Kamnitsas K, et al. Anatomically Constrained Neural Networks (ACNNs): application to Cardiac Image Enhancement and Segmentation. IEEE Trans Med Imaging 2018;37:384–95. doi:10.1109/TMI.2017.2743464.