



RESEARCH ARTICLE

Muscle Mechanics and Ventricular Function

Deep learning-based automated left ventricular ejection fraction assessment using 2-D echocardiography

Xin Liu,^{1*} Yiting Fan,^{2,3*} Shuang Li,⁴ Meixiang Chen,⁵ Ming Li,⁶ William Kongto Hau,⁷  Heye Zhang,⁸ Lin Xu,⁵ and  Alex Pui-Wai Lee^{3,7}

¹Guangdong Academy Research on VR Industry, Foshan University, Guangdong, People's Republic of China; ²Department of Cardiology, Shanghai Chest Hospital, Shanghai JiaoTong University, Shanghai, People's Republic of China; ³Laboratory of Cardiac Imaging and 3D Printing, Li Ka Shing Institute of Health Science, The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China; ⁴General Hospital of the Southern Theatre Command, PLA and Guangdong University of Technology, Guangdong, People's Republic of China; ⁵General Hospital of the Southern Theatre Command, PLA and The First School of Clinical Medicine, Southern Medical University, Guangdong, People's Republic of China; ⁶Faculty of Medicine, Imperial College London, National Heart and Lung Institute, Imperial College London, London, United Kingdom; ⁷Division of Cardiology, Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China; and ⁸School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, People's Republic of China

Abstract

Deep learning (DL) has been applied for automatic left ventricle (LV) ejection fraction (EF) measurement, but the diagnostic performance was rarely evaluated for various phenotypes of heart disease. This study aims to evaluate a new DL algorithm for automated LVEF measurement using two-dimensional echocardiography (2DE) images collected from three centers. The impact of three ultrasound machines and three phenotypes of heart diseases on the automatic LVEF measurement was evaluated. Using 36890 frames of 2DE from 340 patients, we developed a DL algorithm based on U-Net (DPS-Net) and the biplane Simpson's method was applied for LVEF calculation. Results showed a high performance in LV segmentation and LVEF measurement across phenotypes and echo systems by using DPS-Net. Good performance was obtained for LV segmentation when DPS-Net was tested on the CAMUS data set (Dice coefficient of 0.932 and 0.928 for ED and ES). Better performance of LV segmentation in study-wise evaluation was observed by comparing the DPS-Net v2 to the EchoNet-dynamic algorithm ($P = 0.008$). DPS-Net was associated with high correlations and good agreements for the LVEF measurement. High diagnostic performance was obtained that the area under receiver operator characteristic curve was 0.974, 0.948, 0.968, and 0.972 for normal hearts and disease phenotypes including atrial fibrillation, hypertrophic cardiomyopathy, dilated cardiomyopathy, respectively. High performance was obtained by using DPS-Net in LV detection and LVEF measurement for heart failure with several phenotypes. High performance was observed in a large-scale dataset, suggesting that the DPS-Net was highly adaptive across different echocardiographic systems.

NEW & NOTEWORTHY A new strategy of feature extraction and fusion could enhance the accuracy of automatic LVEF assessment based on multiview 2-D echocardiographic sequences. High diagnostic performance for the determination of heart failure was obtained by using DPS-Net in cases with different phenotypes of heart diseases. High performance for left ventricle segmentation was obtained by using DPS-Net, suggesting the potential for a wider range of application in the interpretation of 2DE images.

deep learning; echocardiography; ventricular ejection fraction

INTRODUCTION

The left ventricular ejection fraction (LVEF) is the central measure of left ventricle (LV) systolic function. Although

various methods have been used for the assessment of LVEF, echocardiography remains the first-line approach for its wide availability. In addition, with the advent of handheld ultrasound devices, assessment of LVEF is increasingly done

* X. Liu and Y. Fan contributed equally to this work.

Correspondence: L. Xu (lynne1111@126.com); A. P.-W. Lee (alexpwlee@cuhk.edu.hk).

Submitted 3 June 2020 / Revised 18 June 2021 / Accepted 18 June 2021



at the point of care by healthcare professionals with variable amount of training in echocardiography (1, 2). Manual identification and tracking of the endocardial borders in different views at appropriate frames [end diastole (ED) and end systole (ES)] is expertise dependent, time consuming, and subject to high observer variability (1–5). As a result, the accuracy of LVEF assessment could be suboptimal.

Automatic approaches for LVEF assessment based on two-dimensional echocardiogram (2DE) have been recently introduced (5). Ability to save time and improved reproducibility of the automatic approaches compared with the manual approach has been reported in the previous studies (6–9). But the widespread use of the automated LVEF assessment is still limited by several factors: 1) most commercially available software remain semiautomatically requiring manual initiation of myocardial border tracing by location of anatomic landmarks; 2) implementation remains largely vendor-specific; and 3) accuracy is reduced when image quality is suboptimal. More recently, deep-learning (DL) algorithms were developed to address the clinical needs of automatic view classification (10), fully automatic workflow of heart diseases prediction (11), and high performance on echocardiograms with degraded resolution (12, 13). High performance for view classification (91.7% in accuracy) and diseased predictions with high reproducibility (5, 12) were reported. In addition, data sets used to develop DL algorithms are openly accessible, such as the CAMUS data set (Cardiac Acquisitions for Multistructure Ultrasound Segmentation) (9) and the EchoNet-Dynamic data set (13). However, clinical implication of the DL methods is difficult for its reduced accuracy in rhythm and structural variations, including LV hypertrophy, dilatation, regional wall motion abnormalities, or irregular cardiac cycles (14). Furthermore, a considerable number of data sets with large-scale variations were required to develop a robust DL algorithm fully capable of the tasks of routine clinical usage. Although data sets of various diseases were still resource limited, the quest for a more efficient learning scheme of the deep-learning architecture is of great interest.

In the present study, a newly developed DL algorithm, namely DPS-Net, for automatic LVEF assessment based on sequential 2DE images was evaluated. The DPS-Net was a convolution neural network (CNN) based on a modified U-net with symmetric architecture, targeted on fast and accurate semantic segmentation. Evaluation was performed on the data sets with different heart disease phenotypes and open-source data sets.

METHODS

Data set

There were three data sets used in the present study, including two open-source data sets and one clinical data set collected from our institutions. The two open-source data sets were CAMUS data sets and EchoNet-Dynamic data sets. The CAMUS data set consisted of 500 2DE studies optimized for LVEF calculation collected from one center using the same ultrasound system. Each 2DE study consisted of one apical four-chamber view (A4C) and one apical two-chamber view (A2C). A wide variety of acquisition settings and image

quality was included in the data sets to maintain clinical realism. The EchoNet-Dynamic data set consisted of 10,030 2DE grayscale videos of the A4C collected from 10,030 individuals, using multiple vendors/models of ultrasound systems. All data sets were stored in the Digital Imaging and Communications in Medicine (DICOM) format. Identifying information and human labels were removed from the data sets.

Our data set comprised two modules. *Module 1* was used for algorithm development, and *module 2* was additional data for testing generalization of the algorithm. Three cardiovascular centers [Prince of Wales Hospital (PWH); Third Peoples' Hospital of Shenzhen (TPH); and General Hospital of Southern Theater Command of PLA (STCGH)] participated in this study. Echocardiography was performed using three different ultrasound systems from two vendors, including iE33/EPIQ7C (Philips, Andover, MA) and VIVID E9 (General Electric, Milwaukee, WI). All data sets were anonymized and saved in DICOM format. Electrocardiography-gating data from each study was used to assist the separation of one heartbeat to another in each sequential image. Contrast echocardiography was not used. The *module 1* data set consisted of 100 studies retrospectively collected from 100 individuals undergoing clinically indicated echocardiography: the data were randomly split in 60:20:20 ratio as training (60 studies with 6,515 frames), validation (20 studies with 2,172 frames), and the test (20 studies with 2,171 frames) sets. In addition, the apical three chamber view (A3C) was used to extend the sample size of the data set for DL algorithm development. Thus, the sequential images of these three views were extracted from each study to construct the *module 1* data set. The *module 2* data set consisted of 240 studies including normal hearts and disease phenotypes unseen by the DPS-Net [hypertrophic cardiomyopathy (HCM); dilated cardiomyopathy (DCM); and atrial fibrillation (AF)]. Similarly, the A4C, A3C, and A2C were extracted from each study. Therefore, a total of 36,890 frames of 2DE were included in our local data sets. All data was anonymized. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and received approval from the Institutional Review Boards.

DPS-Net Development and Training

The architecture of the DPS-Net was previously described (15). The structure of the algorithm was constructed based on the U-net, with proven high efficiency and accuracy in medical image segmentation. Four components were included in the architecture for the task of LV segmentation in this study. First, the architecture was a full convolution network (FCN) consisting of five levels in both phases of contraction and expansion networks. In the present study, a dilated dense block was introduced to each level of the contraction path. The dilated dense block was a CNN architecture that concatenated all inputs to the outputs to improve the feature reuse, which has been used for medical image classification with more efficient learning (15). Moreover, plain full convolution (FC) was replaced with a dilated convolution in the dilated dense block to meet the need for large reception and high resolution in the expansion path.

Second, deep supervision was applied by jointing the feature map output from each level of the dilated dense block to each level of the expansion path. The scheme of skip connection between each level of contraction and expansion enables the fusion of the low-level and high-level features. This scheme allowed the network to conduct the semantic segmentation with constraining from the salient features of the LV chamber in the original 2DE image.

Third, a feature pyramid network (FPN) was used to connect the contraction and expansion paths. Four parallel convolutions were adopted in the FPN. Four different scales of receptive fields were used to generate four representative feature maps. These feature maps were concatenated into the same resolution to preserve a large reception field and the subtle variations to enhance the global contextual information of the LV endocardial border.

Finally, the segmentation of LV refined using the predictions of LV regions from five levels of the expansion. The ES and ED frame was automatically extracted from A4C and A2C views, which was indicated as the smallest and largest area in each sequential image from the A4C and A2C views. The volumetric and LVEF calculation was automatically conducted following the biplane Simpson's method based on the frame-based segmentations. In brief, a straight line was traced by connecting the mitral valve level from the lateral to medial annulus points. The height of the cavity (L) is taken from the middle point of the straight line across the mitral valve annulus to the most distal point of the LV apex. The volume of the LV was then calculated as the summation of the disk volumes with an equally discretized height of the cavity ($h = L/n$). The equation was as follows (Eq. 1):

$$\text{Volume (mL)} = \pi \left(\frac{D1}{2} \right) \times \left(\frac{D2}{2} \right) \times h, \quad (1)$$

where $D1$ and $D2$ was the length and width of each disk calculated from A4C and A2C, respectively.

All images were resampled to 256 by 256 and Stochastic Gradient Descent (SDG) with momentum 0.9 was applied to minimize the composite error of LV segmentation. The initial learning rate was 0.002, decay by the coefficient of 0.5 every 50 epochs. The 10-fold cross-validation was used during the training phase. All computations were conducted using a workstation with Xeon 2.1GHz CPU and one NVIDIA Titan XP GPU under Ubuntu 18.04.

Conventional Left Ventricular Segmentation and LVEF Assessment

The endocardial border at A2C, A3C, and A4C views in 36,890 images from all sequential images were manually delineated by two experienced echocardiographers (A.L., Y. F.). The LVEF was assessed based on the A4C and A2C using the Simpson's method (16). The manual tracing of the LV border and calculated LVEF were considered as reference for the assessment of the automatic method.

Performance Assessment of DPS-Net for Automatic Volumetric and LVEF Calculation

We evaluated the algorithm in the following aspects. First, the performance of the DPS-Net was evaluated by comparing the LV segmentations between manual and automatic

methods. The accuracy of the automatic LV segmentation was evaluated by using the Dice coefficient, which was defined as the portion of the overlapped area between the predicted LV chamber and the manual tracing LV chamber in each frame of the sequential 2DE images. The Dice coefficient was calculated as follows (Eq. 2):

$$\text{Dice} = \frac{2 \times (A_{DL} \text{ within } A_g)}{(A_{DL} + A_g)}, \quad (2)$$

where A_{DL} was area of LV predicted by DL algorithm and A_g was ground truth area of LV calculated from manual tracing border. The overall performance of DL algorithm for LVEF calculation was evaluated as well as several subgroups according to different echocardiography systems and disease phenotypes. In addition, automatic LVEF measurement was evaluated by using the following strategies: 1) ED and ES frames extracted from a pair of selected cycle in A2C and A4C were used to calculate LVEF, 2) three continuous cycles in A2C and A4C were paired in the order of appearance that final LVEF was the averaged of LVEF from the three cycles.

Second, the inter- and intraobserver variability analysis was conducted. A subgroup data set consisted of 50 cases randomly selected from the *module 2* data sets, named as observer variation data set (OV data set). Manual classification of the ED/ES frames and segmentation of the LV was performed by two specialists and repeated 7 days later on the same machine. The LV segmentation and the ensuring LVEF calculations were compared. The repeated calculation was also performed by using the DPS-Net on the sub-data set. The inter- and intraobserver variability were evaluated by intraclass correlation coefficients (ICCs).

Third, the CAMUS data set and the EchoNet-dynamic data set were used as external data sets to evaluate the DPS-Net. However, there is a lack of frame-based label in the public data source. The frame-by-frame labeling on the public data sets was performed by our heart-team specialists of 22 individuals for more comprehensive evaluation. A total of 10,030 studies of A4C view from EchoNet-dynamic data set were labeled frame-by-frame. The labels were further reviewed by three senior experts. Recognizing the variation from image preparation between our data set and the public data set, we used the DPS-Net as a pretrain model to further calibrate the weights by inheriting the hyperparameter settings of DPS-Net using the 90% of the EchoNet-dynamic data set, named as DPS-Net v2. A 10-fold validation procedure was performed. Evaluation was performed by using the rest of the EchoNet-dynamic data set (10% of EchoNet-dynamic data set) and CAMUS data set. The frame-based Dice similarity coefficient was used to evaluate the DPS-Net v2 on the external data sets. Comparison between the DPS-Net v2 and the EchoNet-dynamic algorithm was also conducted by evaluating the LV segmentation on the EchoNet-dynamic data set and the *module 2* data set from our study.

Statistical Analysis

Data analysis was performed using SPSS 24.0 (SPSS Inc). Continuous variables are presented as means \pm SD, unless otherwise specified. Categorical variables are presented as

Table 1. Characteristics of local data sets

	Module 1 Data set	Module 2 Data set
<i>n</i>	100	240
Age, yr	64 ± 15	63 ± 11
Men, <i>n</i> (%)	79 (79.0)	151 (62.9)
Systolic blood pressure, mmHg	134 ± 21	132 ± 22
Diastolic blood pressure, mmHg	75 ± 14	75 ± 13
Height, cm	162 ± 11	161 ± 13
Weight, kg	65 ± 14	64 ± 13
Hypertension, <i>n</i> (%)	79 (79.0)	182 (75.8)
Diabetes mellitus, <i>n</i> (%)	45 (45.0)	97 (40.4)
Chronic kidney diseases, <i>n</i> (%)	34 (34.0)	20 (8.3)
Coronary artery disease, <i>n</i> (%)	53 (53.0)	110 (45.8)
Atrial fibrillation, <i>n</i> (%)		49 (20.4)
Hypertrophy cardiomyopathy, <i>n</i> (%)		25 (10.4)
Dilate cardiomyopathy, <i>n</i> (%)		25 (10.4)
Reduced EF (≤50%), <i>n</i> (%)	37 (37.0)	140 (58.3)
Ultrasound system, <i>n</i> (%)		
EPIQ 7 C	50 (50.0)	191 (79.6)
iE33	22 (22.0)	33 (13.8)
Vivid E9	28 (28.0)	16 (6.7)

Values are means ± SD or *n* (%). Baseline characteristics of patients. EF, ejection fraction.

numbers and percentages. Group comparisons were conducted with one-way ANOVA test or the Mann–Whitney *U* test as appropriate. Correlation analysis between LVEF by manual tracking and DL algorithm was performed by using the Pearson correlation. Diagnostic performance was also evaluated by accuracy, specificity, sensitivity, PPV, and NPV to classify LVEF as reduced (<50%) or preserved (>50%).

RESULTS

Table 1 summarized the characteristics of the study population in local data sets (*modules 1* and *2*). The average time for calculating one case was 15.2 ± 8.1 s/patient for segmentation of ventricular border and 8 ± 3.3 s/patient for generating calculations of LV volume and LVEF using automated analysis.

Performance of DPS-Net in LV Segmentation on Local Data sets

The Dice coefficient calculated from all frames in each view of the echocardiographic studies summarized in Table

2. Frame-based Dice similarity coefficients of all frames were >0.9 for all views. No significant difference in identifying ED/ES frame for each view; performance was consistent across different disease phenotypes, although a slightly lower Dice coefficient was found in patients with AF. Model-based analysis showed high performance in LV segmentation regardless of ultrasound systems from *module 2* data set (Fig. 1).

Generalization of the Functional Assessment to Different Phenotypes

Although using a single beat for the automatic LVEF measurement, the accuracy, sensitivity, specificity, positive prediction value, and negative prediction values of the DPS-Net for automatic LVEF (<50 vs. >50%) in *module 2* data set was 90%, 91%, 89%, 86%, and 93%, respectively. The area under the receiver operating characteristic (ROC) curve was 0.974 for classification of reduced heart function. High correlation and good agreement were found when automatic LVEF was calculated from single cycle using selected pair of A2C and A4C (Fig. 2). Homoscedasticity of the automatic LVEF calculation was observed in the cases of AF (*P* value = 0.299), HCM (*P* value = 0.697), and DCM (*P* value = 0.746) compared with the result of the overall data set. When the automatic LVEF was calculated by averaging multiple cycles, automatic LVEF measurement showed a better performance that higher correlation and smaller limited of the agreement (Fig. 3).

Evaluation of the Inter- and Intraobserver Variation

Repeated automatic LVEF measurement using DPS-Net on the OV data set resulted in the highest level of consistency (ICC = 0.998), which is better than the accuracy of clinical specialists' measurement. A separate comparison for the individual LVEF measurement showed that our DL algorithm is better than the repeat measurements of the specialists (Table 3). Our DL algorithm showed the least mean absolute error (MAE) and root mean standard error (RMSE) of 1.2% and 1.3% for automatic LVEF measurement. The MAE and RMSE of the two specialists for repeated LVEF measurement was 3.5% ± 2.7% and 4.9% ± 4.0%, respectively.

Table 2. Performance of the DPS-Net for segmentation in all frames

	A2C	A3C	A4C	<i>P</i> Value
Study-based analysis				
Overall	0.931 [0.928–0.935]	0.931 [0.922–0.939]	0.933 [0.924–0.942]	0.89
ED	0.942 [0.938–0.945]	0.933 [0.931,0.936]	0.930 [0.928,0.932]	0.52
ES	0.928 [0.923,0.934]	0.919 [0.915,0.923]	0.922 [0.920,0.924]	0.55
Phenotype-based analysis				
EF < 50	0.930 [0.929,0.931]	0.930 [0.929,0.932]	0.929 [0.927,0.931]	0.55
EF > 50	0.940 [0.939,0.941]	0.942 [0.941,0.944]	0.939 [0.936,0.943]	0.79
AF	0.922 [0.918,0.925]	0.920 [0.912,0.929]	0.920 [0.910,0.930]	0.69
HCM	0.939 [0.937,0.940]	0.931 [0.928,0.933]	0.931 [0.930,0.932]	0.72
DCM	0.928 [0.927,0.930]	0.925 [0.921,0.929]	0.930 [0.929,0.932]	0.69
Machine-based analysis				
Epiq 7 C	0.933 [0.932,0.935]	0.932 [0.930,0.934]	0.944 [0.940,0.948]	0.48
iE33	0.929 [0.925,0.934]	0.930 [0.921,0.939]	0.935 [0.928,0.943]	0.75
Vivid 9	0.931 [0.923,0.939]	0.928 [0.924,0.932]	0.932 [0.927,0.937]	0.71

Values are Dice coefficients [95% confidence intervals]. A2C, A3C, A4C, apical 2-, 3-, or 4-chamber views, respectively; AF, atrial fibrillation; ED, end diastole; ES, end systole; EF, ejection fraction; HCM, hypertrophic cardiomyopathy; DCM, dilated cardiomyopathy.

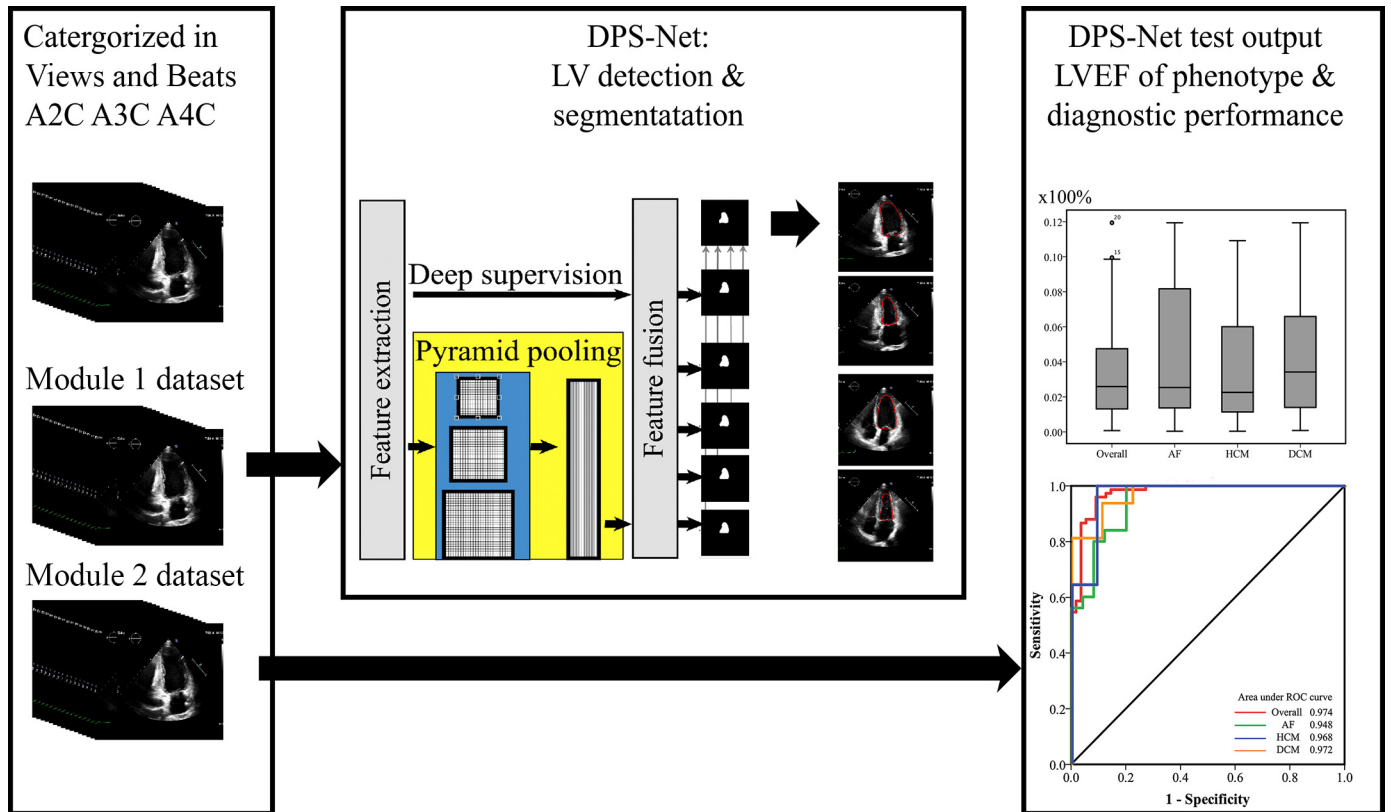


Figure 1. Model development and evaluations. *Module 1* data set ($n=100$) was used to develop DPS-Net and the *module 2* data set was used to evaluate the performance of DPS-Net. Skip connection between feature extraction and fusion path acts as deep supervision by using the low-level feature to supervise high-level semantic segmentation. Pyramid pooling modular combines 3-dimensional features to preserve the detail of the images. Predictions of the LV area from each layer of the fusion path was used to predict the final output of the LV segmentation. The low mean absolute error was observed in all cases and cases with HCM and DCM, particularly. High diagnostic performance was obtained in each phenotype. The area under ROC curve was 0.974, 0.968, 0.972, and 0.948 in all cases, cases of HCM, DCM, and AF. AF, atrial fibrillation; DCM, dilated cardiomyopathy; EF, ejection fraction; HCM, hypertrophic cardiomyopathy; LVEF, left ventricular ejection fraction.

Performance of DPS-Net in LV Segmentation on External Datasets

The evaluation in the CAMUS data set showed high Dice coefficients in LV segmentation by DPS-Net V2 with 0.932 [95% CI: (0.938,0.925)] and 0.928 [95% CI: (0.924,0.932)] for ED and ES frames, respectively. Our automatic LVEF measurement was in good agreement with the reference standard, reflected by a means \pm SD of $-1.70\% \pm 4.13\%$ with 95% CI ($-9.79, 6.39$). This was better than the accuracy from the previous study in that the means \pm SD was $1.8\% \pm 8.9\%$ (17). Table 4 summarized the evaluation of LV segmentation in the EchoNet-dynamic data sets. DPS-Net v2 was comparable or slightly better than the accuracy of the EchoNet-dynamic algorithm when testing in the EchoNet-dynamic data set. Evaluation of the LV segmentation in the *module 2* data set showed that higher accuracy was obtained by DPS-Net v2 compared with the EchoNet-dynamic algorithm ($P < 0.001$).

DISCUSSION

In this study, we evaluated a newly developed deep learning algorithm for LV segmentation on echocardiograms from patients with different phenotypes of heart diseases (18). Our results showed that DPS-Net was associated with high

accuracy in LV segmentation regardless of rhythmic and structural variation on the 2DE images. High reproducibility and consistency were obtained by our DL algorithm according to the high Dice coefficients. This improvement of LV anatomic assessment indicated the potential of DPS-Net for screening of heart conditions in routine clinical application.

In terms of echocardiographic interpretation, previous studies reported that automatic methods may be a benefit to the early recognition of heart diseases (8, 11, 13, 19). As anatomical and rhythmic variation are the two major abnormalities of the cardiac dysfunction, LVEF alone is not an efficient indicator to HCM as persevered LVEF is commonly found (20, 21), whereas the LV ED volumes or diameters $> 2SD$ was the criteria used for the identification of DCM (18). Zhang et al. demonstrated the capability of deep learning-based automatic segmentation of the myocardial chamber and heart wall for classifying different phenotypes of cardiac dysfunction with distinguishing distribution of the chamber volume and mass (11). Kusunose et al. showed that a deep-learning algorithm could further classify the heart dysfunction with regional wall motion abnormalities (19). Despite of these findings, the accuracy of the LV border detection in different phenotypes has been rarely examined, resulting in the gap between image processing and functional interpretation in the previous studies (8, 11, 14). In this study, our frame-based

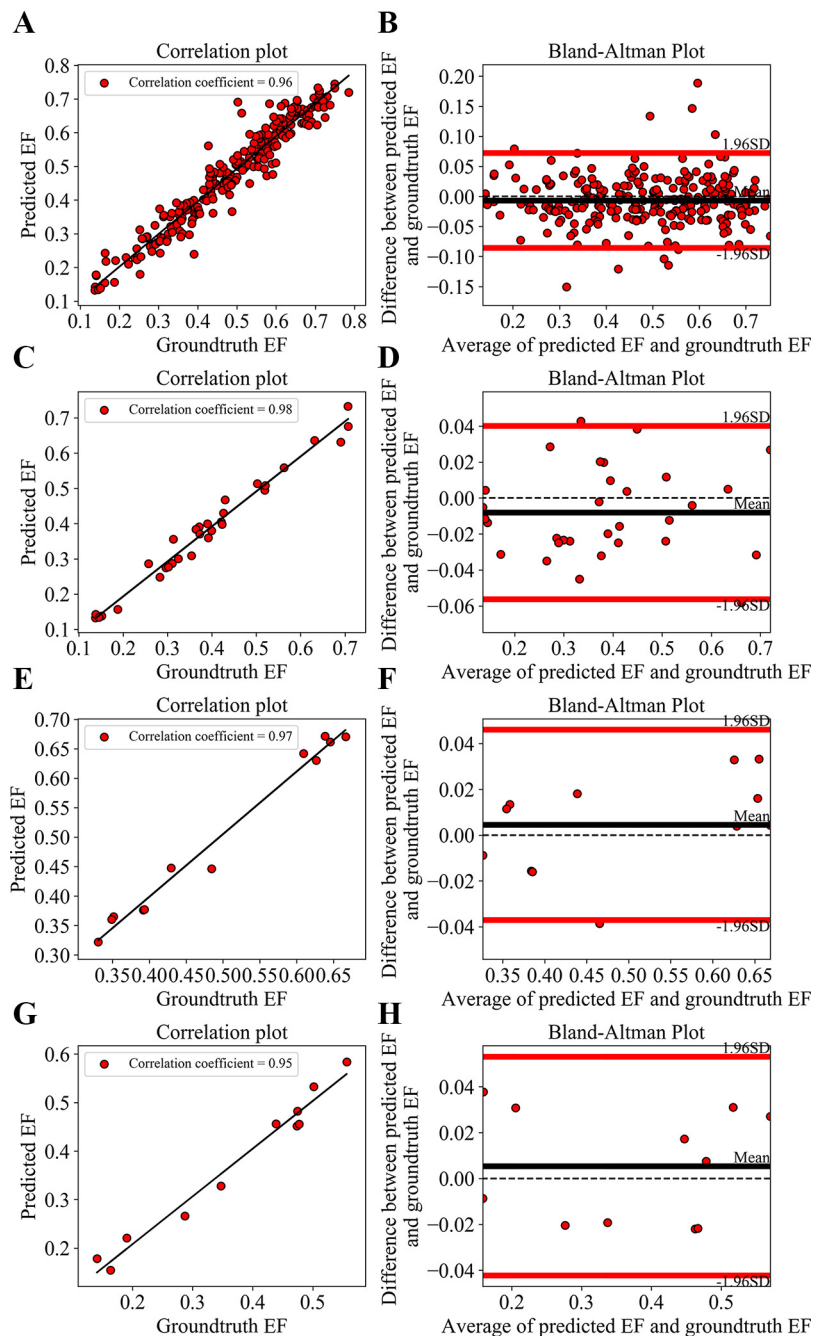


Figure 2. High correlation and good agreement in LVEF calculation were achieved by the DPS-Net by taking the manual LVEF as reference. Correlation coefficient with the limit of agreement (LOA) was 0.96 with $[-0.09, 0.07]$, 0.94 with $[-0.11, 0.07]$, 0.94 with $[-0.10, 0.08]$, and 0.92 with $[-0.10, 0.07]$ for overall performance (A and B), cases with AF (C and D), HCM (E and F), and DCM (G and H), respectively. AF, atrial fibrillation; DCM, dilated cardiomyopathy; EF, ejection fraction; HCM, hypertrophic cardiomyopathy; LVEF, left ventricular ejection fraction.

deep-learning algorithm also showed consistent accuracy over the sequential 2DE images from patients with different phenotypes, which enabled the more precise automatic assessment of LV anatomy. Thus, the DPS-Net was capable to provide a more objective functional interpretation based on 2DE.

The generalization of the DL algorithm is crucial in the extension of clinical applications. Given the unstructured and imbalanced nature of the medical data set, it remains a challenge to generalize DL algorithms for the increasing data sets unmet previously. Different settings of image acquisition could degrade the accuracy of echocardiogram interpretation (22). Despite the variation of the image quality remained a

significant concern on the diagnostic accuracy (23, 24), our fine-tuned DPS-Net V2 demonstrated a better performance in LV segmentation on the 2DE image from external and our *module 2* data set compared with the previous work (13). By learning about the relationship between data and corresponding labels (25), the performance of the deep-learning model varied and relied on the feature construction in the intermediate layers (19). The novel DL algorithm tested in this study reflected that two aspects contributed the most to the superior performance. First, the “deep supervision” and pyramid pooling modules were applied to modify the original U-Net in the proposed DL algorithm. Deep supervision was architecture as a skip-connection

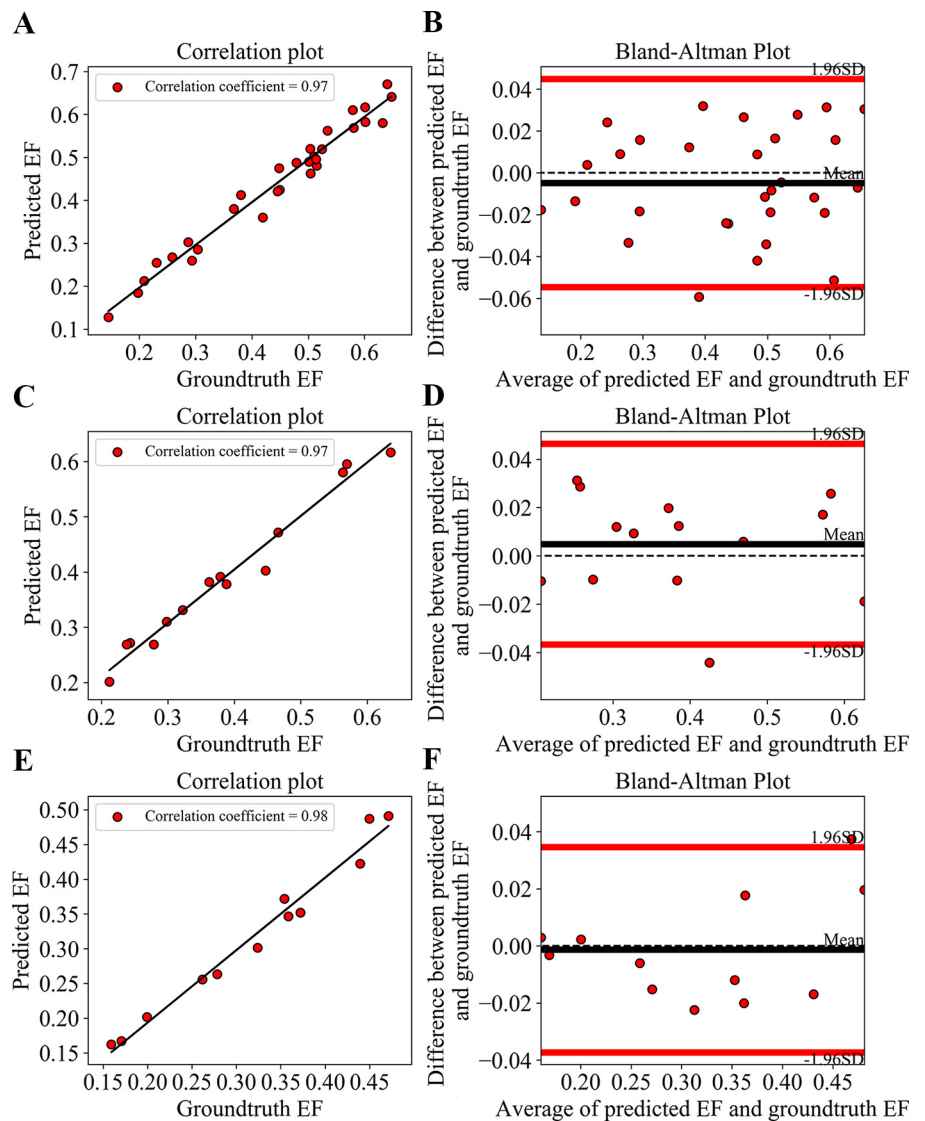


Figure 3. Better performance was found when using averaging of multiple cycles for automatic LVEF measurement in phenotypes. Correlation coefficient with the LOA was 0.97 with $[-0.05, 0.04]$, 0.97 with $[-0.04, 0.05]$, and 0.98 with $[-0.04, 0.03]$ for cases with AF (A and B), HCM (C and D), and DCM (E and F), respectively. AF, atrial fibrillation; DCM, dilated cardiomyopathy; EF, ejection fraction; HCM, hypertrophic cardiomyopathy; LVEF, left ventricular ejection fraction.

between low-level features and high-level semantic features, which allowed a better reconstruction of the LV area. A previous study demonstrated the improvement of view classification and endocardial border segmentation on standard 2DE images and by implementing skip connection (26, 27). Second, the pyramid pooling modular was used to translate the key features in four dimensions convolution blocks as a “pyramid” shape. This modular preserved the gradient of pixels and their relationship to adjacent pixels to the local and global features. As a result,

a more accurate endocardial border segmentation led to a better LVEF prediction compared with previous work (13).

Study Limitations

First, the modality used in this study was limited. On one hand, we developed the DL algorithm for LV assessment based on 2DE when the recent chamber quantification guidelines emphasized that 3DE measurements are preferable for superior accuracy and reproducibility (16, 28). However, 2DE is more commonly used for LV assessment

Table 3. Inter- and intraobserver variation analysis

	First Time			
	Reference	Specialist 1	Specialist 2	DPS-Net
Repeated				
Reference		0.940 [0.878,0.968]	0.956 [0.917,0.976]	0.969 [0.965,0.976]
Specialist 1	0.948 [0.891,0.973]	0.961 [0.931,0.978]	0.957 [0.925,0.976]	
Specialist 2	0.945 [0.890,0.970]	0.964 [0.937,0.980]	0.953 [0.917,0.973]	
DPS-Net	0.969 [0.965,0.976]			0.998 [0.997,0.999]

Values are intraclass correlations [95% confidence intervals].

Table 4. Comparison of the performance for left ventricular detection

	EchoNet-Dynamic Data set			Module 2 Data set		
	Study-Wise	ED	ES	Study-Wise	ED (A4C)	ES (A4C)
DPS-Net V2	0.925 [0.915,0.936]	0.925 [0.920,0.929]	0.923 [0.914,0.931]	0.935 [0.929,0.940]	0.934 [0.925,0.943]	0.929 [0.923,0.935]
EchoNet-dynamic	0.910 [0.901,0.919]	0.925 [0.922,0.928]	0.916 [0.914,0.918]	0.907 [0.898,0.916]	0.922 [0.919,0.924]	0.910 [0.906,0.915]
P value	0.008	0.81	0.79	<0.001	0.12	0.04

Values are Dice coefficients [95% confidence intervals]. A4C, apical 4-chamber views; ED, end diastole; ES, end systole.

that our DL algorithm for 2DE analysis is of high practical value (29, 30). On the other hand, a previous study reported the safety and accuracy of contrast echocardiography for LV functional assessment (31). By using the framework of U-net, previous work demonstrated the feasibility of the DL algorithm for myocardial wall detection based on myocardial contrast echocardiographic image (32). Recognizing the similarity of our DL algorithm to the previous work, adequate tuning of the parameters should the proposed algorithm be adjusted to the data set of myocardial contrast echocardiographic image in future work.

Second, several factors that impact the accuracy of the biplane 2DE interpretation were not included in the present study, including beat-to-beat variations, foreshortening, and poor image quality. For the beat-to-beat variations, the dyssynchronization between A2C and A4C views in cases with AF could introduce additional error into the LV assessment. Although biplane Simpson's method was preferable for LVEF measurement, averaging from additional cardiac cycles was recommended to avoid those beats with significantly shortened cycle length (16, 33). As the ED and ES frames were automatically identified according to the largest and smallest areas from the LV segmentation phase of the DL algorithm, the selected pair of A2C and A4C contained a single cycle with the largest area of ED frames. Automatic LVEF measurement was therefore calculated from a relatively normal cardiac cycle. When automatic LVEF was averaged from multiple cardiac cycles, ED and ES frames were recorded from three complete cycles and paired in the order of appearance in each view to generate LVEF. Random pairing was not included. However, our result demonstrated that the DPS-Net was associated with a high correlation to the reference in both strategies. With the improvement of our DL algorithm for 2DE assessment, further evaluation in cases with significant beat-to-beat variations could be conducted when image acquisition fulfills the clinical guidelines. For the foreshortening, there was a detection and calibration method based on the maximum axial distance that apical foreshortening could be alleviated (17). For the variation of image qualities, previous studies illustrated the utilities for guidance on the acquiring of a diagnostic echocardiogram by using support from machine learning algorithms (34, 35). The accurate acquisition of the plane of views could significantly increase the effectiveness of the LVEF assessment. The present study focused on the automatic interpretation of the 2DE images, the guidance of the image acquisition process will be integrated to meet the needs of real-world clinical practice.

Third, the sample size used in the present study was small. The smaller sizes of data sets from less commonly used echocardiographic systems could be of insufficient statistical

power. However, our algorithm illustrated an excellent performance for detection of the myocardial border on large size publicly available datasets (13). Therefore, our algorithm demonstrated a superior performance for learning the semantic features of the LV anatomic from small volume of data set. Alternative machine-learning-based approximation based on dimension-less time-dependent contraction coefficient was developed for LVEF prediction to reduce the cumulative variation from LV segmentation (36). Further study will be conducted to compare the diagnostic performance of these methods.

Fourth, automatic view classification was not included in the present study. Previous work emphasized that view classification was a key component of the translation of the DL algorithm into real-world clinical (10). However, automatic LVEF measurement for a wide range of phenotypes remained a challenge in 2DE interpretation (13, 19). Further studies are required for a more comprehensive evaluation of the deep learning-based systems on interpreting heart function based on multiple views of echocardiographic images.

Conclusions

In conclusion, we evaluated a more efficient DL algorithm for automatic LVEF measurement based on 2DE. Based on the high performance of LV segmentation, DPS-Net showed superior accuracy in generating LVEF across several phenotypes. In addition, the evaluation of the combination of our data set and large-scale external data set suggested that the DPS-Net was highly adaptive across different echocardiographic systems.

SUPPLEMENTAL DATA

Supplemental Material: <https://github.com/Heye-SYSU/DPS-Net>.

GRANTS

This study was funded by Natural Science Foundation (Guangdong Province, China) Grants 2019A151011463, 2018A050506031, 2017B010125001 and 2019B010110001; Hong Kong Special Administrative Region Health and Medical Research Fund Grant 05160976; National Natural Science Foundation of China, International (regional) Cooperation and Exchange Grant 81911530252; and National Natural Science Foundation of China Grants 61771464, U1801265, 82001910, and 61976222.

DISCLOSURES

Prof. Alex Lee is on the speakers' bureau of Philips Healthcare. None of the other authors has any conflicts of interest, financial or otherwise, to disclose.

AUTHOR CONTRIBUTIONS

X.L., W.K.H., H.Z., L.X., and A.P.-W.L. conceived and designed research; X.L., Y.F., S.L., M.C., and M.L. performed experiments; X.L., Y.F., S.L., M.C., and M.L. analyzed data; X.L., Y.F., S.L., M.C., and M.L. interpreted results of experiments; X.L. prepared figures; X.L. drafted manuscript; X.L., H.Z., and A.P.-W.L. edited and revised manuscript; X.L., Y.F., S.L., M.C., M.L., W.K.H., H.Z., L.X., and A.P.-W.L. approved final version of manuscript.

REFERENCES

- Sugimoto T, Dulgheru R, Bernard A, Ilardi F, Contu L, Addetia K, et al. Echocardiographic reference ranges for normal left ventricular 2D strain: results from the EACVI NORRE study. *Eur Heart J Cardiovasc Imaging* 18: 833–840, 2017. doi:10.1093/ehjci/jex140.
- Yasufumi N, Yuichiro K, Takeshi O, Kyoko O, Akemi N, Yutaka O, Masaaki T. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. *Echo Res Pract* 5: 27–39, 2018. doi:10.1530/ERP-17-0047.
- Cameli M, Mondillo S, Solari M, Righini FM, Andrei V, Contaldi C, De Marco E, Di Mauro M, Esposito R, Gallina S, Montisci R, Rossi A, Galderisi M, Nistri S, Agricola E, Mele D. Echocardiographic assessment of left ventricular systolic function: from ejection fraction to torsion. *Heart Fail Rev* 21: 77–94, 2016. doi:10.1007/s10741-015-9521-8.
- Guo Y, Green S, Park L, Rispen L. Left ventricle volume measuring using echocardiography sequences. *2018 Digital Image Computing: Techniques and Applications (DICTA)* 2018, p. 1–8. doi:10.1109/DICTA.2018.8615766.
- Nolan MT, Thavendiranathan P. Automated quantification in echocardiography. *JACC Cardiovasc Imaging* 12: 1073–1092, 2019. doi:10.1016/j.jcmg.2018.11.038.
- Abazid RM, Abohamr SI, Smettei OA, Gasem MS, Suresh AR, Al Harbi MF, Aljaber AN, Al Motairy AA, Albiela DE, Almutairi BM, Sakr H. Visual versus fully automated assessment of left ventricular ejection fraction. *Avicenna J Med* 8: 41–45, 2018. doi:10.4103/ajm.209_17.
- Frederiksen CA, Juhl-Olsen P, Hermansen JF, Andersen NH, Sloth E. Clinical utility of semi-automated estimation of ejection fraction at the point-of-care. *Heart Lung Vessel* 7: 208–216, 2015.
- Knackstedt C, Bekkers SCAM, Schummers G, Schreckenber M, Muraru D, Badano LP, Franke A, Bavishi C, Omar AMS, Sengupta PP. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: The FAST-EFs multicenter study. *J Am Coll Cardiol* 66: 1456–1466, 2015. doi:10.1016/j.jacc.2015.07.052.
- Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin P, Grenier T, Lartizien C, D'hooge J, Lovstakken L, Bernard O. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging* 38: 2198–2210, 2019. doi:10.1109/TMI.2019.2900516.
- Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digital Med* 1: 6, 2018. doi:10.1038/s41746-017-0013-1.
- Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation* 138: 1623–1635, 2018. doi:10.1161/CIRCULATIONAHA.118.034338.
- Hu Y, Xia B, Mao M, Jin Z, Du J, Guo L, Frangi AF, Lei B, Wang T. AIDAN: an attention-guided dual-path network for pediatric echocardiography segmentation. *IEEE Access* 2020, 8, pp. 29176–29187. doi:10.1109/ACCESS.2020.2971383.
- Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 580: 252–256, 2020. doi:10.1038/s41586-020-2145-8.
- Kusunose K, Haga A, Yamaguchi N, Abe T, Fukuda D, Yamada H, Harada M, Sata M. Deep learning for assessment of left ventricular ejection fraction from echocardiographic images. *J Am Soc Echocardiogr* 33: 632–635, 2020. doi:10.1016/j.echo.2020.01.009.
- Li M, Dong S, Gao Z, Feng C, Xiong H, Zheng W, Ghista D, Zhang H, de Albuquerque VHC. Unified model for interpreting multi-view echocardiographic sequences without temporal information. *Applied Soft Computing* 88: 106049, 2020. doi:10.1016/j.asoc.2019.106049.
- Lang RM, Badano LP, Mor-Avi V, Afzalalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru D, Picard MH, Rietzschel ER, Rudski L, Spencer KT, Tsang W, Voigt J-U. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 16: 233–271, 2015. [Erratum in *Eur Heart J Cardiovasc Imaging* 17(4): 412, 2016]. doi:10.1093/ehjci/jev014.
- Smistad E, Østvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, Brunvand H, Edvardsen T, Leclerc S, Bernard O, Grenne B, Løvestakken L. Real-time automatic ejection fraction and foreshortening detection using deep learning. *IEEE Trans Ultrason Ferroelectr Freq Control* 67: 2595–2604, 2020. doi:10.1109/TUFFC.2020.2981037.
- Schultheiss H-P, Fairweather D, Caforio ALP, Escher F, Hersberger RE, Lipshultz SE, Liu PP, Matsumori A, Mazzanti A, McMurray J, Priori SG. Dilated cardiomyopathy. *Nat Rev Dis Primers* 5: 32, 2019. doi:10.1038/s41572-019-0084-1.
- Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging* 13: 374–381, 2020. doi:10.1016/j.jcmg.2019.02.024.
- Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, Hagege AA, Lafont A, Limongelli G, Mahrholdt H, McKenna WJ, Mogensen J, Nihoyannopoulos P, Nistri S, Pieper PG, Pieske B, Rapezzi C, Rutten FH, Tilmann C, Watkins H; Authors/Task Force members. 2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy: the task force for the diagnosis and management of hypertrophic cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J* 35: 2733–2779, 2014. doi:10.1093/eurheartj/ehu284.
- Haland TF, Hasselberg NE, Almaas VM, Dejgaard LA, Saberniak J, Leren IS, Berge KE, Haugaa KH, Edvardsen T. The systolic paradox in hypertrophic cardiomyopathy. *Open Heart* 4: e000571, 2017. e000571 doi:10.1136/openhrt-2016-000571.
- Luong C, Liao Z, Abdi A, Girgis H, Rohling R, Gin K, Jue J, Yeung D, Szefer E, Thompson D, Tsang MY-C, Lee PK, Nair P, Abolmaesumi P, Tsang TSM. Automated estimation of echocardiogram image quality in hospitalized patients. *Int J Cardiovasc Imaging* 37: 229–239, 2021. doi:10.1007/s10554-020-01981-8.
- Leclerc S, Grenier T, Espinosa F, Bernard O. A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data. *2017 IEEE International Ultrasonics Symposium (IUS)*, 2017, pp. 1–4. doi:10.1109/ULTSYM.2017.8092632.
- Millitari F, Rothberg A, Jia J, Sofka M. Integrating statistical prior knowledge into convolutional neural networks. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, edited by Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S. Cham: Springer International Publishing, 2017, p. 161–168.
- Kwon J-M, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography* 36: 213–218, 2019. doi:10.1111/echo.14220.
- Long G, Benjamin WS. Abstract 276: Identification of Echocardiographic Imaging View Using Deep Learning. *Circ Cardiovasc Qual Outcomes* 11: A276–A276, 2018. doi:10.1161/circoutcomes.11.suppl_1.276.
- Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-Time Standard View Classification in Transthoracic Echocardiography Using Convolutional Neural Networks. *Ultrasound Med Biol* 45: 374–384, 2019. doi:10.1016/j.ultrasmedbio.2018.07.024.
- Medvedofsky D, Maffessanti F, Weinert L, Nehrani DM, Narang A, Addetia K, Mediratta A, Besser SA, Maor E, Patel AR, Spencer KT, Mor-Avi V, Lang RM. 2D and 3D echocardiography-derived indices of left ventricular function and shape: relationship with mortality.

- JACC Cardiovasc Imaging* 11: 1569–1579, 2018. doi:[10.1016/j.jcmg.2017.08.023](https://doi.org/10.1016/j.jcmg.2017.08.023).
29. Medvedofsky D, Mor-Avi V, Amzulescu M, Fernández-Golfín C, Hinojar R, Monaghan MJ, Otani K, Reiken J, Takeuchi M, Tsang W, Vanoverschelde J-L, Indrajith M, Weinert L, Zamorano JL, Lang RM. Three-dimensional echocardiographic quantification of the left heart chambers using an automated adaptive analytics algorithm: multicentre validation study. *Eur Heart J Cardiovasc Imaging* 19: 47–58, 2018. doi:[10.1093/ehjci/jew328](https://doi.org/10.1093/ehjci/jew328).
 30. van den Hoven AT, Mc-Ghie JS, Chelu RG, Duijnhouwer AL, Baggen VJM, Coenen A, Vletter WB, Dijkshoorn ML, van den Bosch AE, Roos-Hesselink JW. Transthoracic 3D echocardiographic left heart chamber quantification in patients with bicuspid aortic valve disease. *Int J Cardiovasc Imaging* 33: 1895–1903, 2017. doi:[10.1007/s10554-017-1192-1](https://doi.org/10.1007/s10554-017-1192-1).
 31. Lindner JR. Contrast echocardiography: current status and future directions. *Heart* 107: 18–24, 2021. doi:[10.1136/heartjnl-2020-316662](https://doi.org/10.1136/heartjnl-2020-316662).
 32. Li M, Zeng D, Xie Q, Xu R, Wang Y, Ma D, Shi Y, Xu X, Huang M, Fei H. A deep learning approach with temporal consistency for automatic myocardial segmentation of quantitative myocardial contrast echocardiography. *Int J Cardiovasc Imaging* 37: 1967–1978, 2021. doi:[10.1007/s10554-021-02181-8](https://doi.org/10.1007/s10554-021-02181-8).
 33. Kim T-S, Youn H-J. Role of echocardiography in atrial fibrillation. *J Cardiovasc Ultrasound* 19: 51–61, 2011. doi:[10.4250/jcu.2011.19.2.51](https://doi.org/10.4250/jcu.2011.19.2.51).
 34. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, Chaudhry A, Martin RP, McCarthy PM, Rubenson DS, Goldstein S, Little SH, Lang RM, Weissman NJ, Thomas JD. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol* 6: 624, 2021. doi:[10.1001/jamacardio.2021.0185](https://doi.org/10.1001/jamacardio.2021.0185).
 35. Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, Gollasch G, Hengstenberg C, Binder T. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. *Int J Cardiovasc Imaging* 37: 577–586, 2021. doi:[10.1007/s10554-020-02046-6](https://doi.org/10.1007/s10554-020-02046-6).
 36. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, Hong H, Mor-Avi V, Martin RP, Lang RM. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circ Cardiovasc Imaging* 12: e009303, 2019. doi:[10.1161/CIRCIMAGING.119.009303](https://doi.org/10.1161/CIRCIMAGING.119.009303).