ESC European Society of Cardiology

European Heart Journal - Digital Health (2024) **5**, 260–269
https://doi.org/10.1093/ehjdh/ztae015

ORIGINAL ARTICLE
Artificial intelligence (machine learning, deep learning)

Artificial intelligence-based classification of echocardiographic views

Jwan A. Naser †, Eunjung Lee†, Sorin V. Pislaru , Gal Tsaban, Jeffrey G. Malins, John I. Jackson, D. M. Anisuzzaman, Behrouz Rostami, Francisco Lopez-Jimenez , Paul A. Friedman, Garvan C. Kane, Patricia A. Pellikka , and Zachi I. Attia *

Department of Cardiovascular Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

Received 20 May 2023; revised 21 February 2024; accepted 22 February 2024; online publish-ahead-of-print 26 February 2024

Aims Augmenting echocardiography with artificial intelligence would allow for automated assessment of routine parameters and identification of disease patterns not easily recognized otherwise. View classification is an essential first step before deep learning can be applied to the echocardiogram.

Methods and results We trained two- and three-dimensional convolutional neural networks (CNNs) using transthoracic echocardiographic (TTE) studies obtained from 909 patients to classify nine view categories (10 269 videos). Transthoracic echocardiographic studies from 229 patients were used in internal validation (2582 videos). Convolutional neural networks were tested on 100 patients with comprehensive TTE studies (where the two examples chosen by CNNs as most likely to represent a view were evaluated) and 408 patients with five view categories obtained via point-of-care ultrasound (POCUS). The overall accuracy of the two-dimensional CNN was 96.8%, and the averaged area under the curve (AUC) was 0.997 on the comprehensive TTE testing set; these numbers were 98.4% and 0.998, respectively, on the POCUS set. For the three-dimensional CNN, the accuracy and AUC were 96.3% and 0.998 for full TTE studies and 95.0% and 0.996 on POCUS videos, respectively. The positive predictive value, which defined correctly identified predicted views, was higher with two-dimensional rather than three-dimensional networks, exceeding 93% in apical, short-axis aortic valve, and parasternal long-axis left ventricle views.

Conclusion An automated view classifier utilizing CNNs was able to classify cardiac views obtained using TTE and POCUS with high accuracy. The view classifier will facilitate the application of deep learning to echocardiography.

Keywords Artificial intelligence • Machine learning • Deep learning • Echocardiography • View classification • Ultrasound • Neural network

Introduction

Artificial intelligence (AI) has shown promise in various fields in medicine.[1–3] Through recognition of complex patterns not easily identifiable by the human eye, not only has deep learning boosted radiological diagnosis of diseases,[4,5] but it has also expanded the utility of inexpensive tests, such as the electrocardiogram (ECG), to detect diseases not otherwise routinely recognized as in the identification of paroxysmal atrial fibrillation during an ECG in sinus rhythm.[6] Transthoracic echocardiography (TTE), on the other hand, gives extensive information on the structure and function of the heart and is widely available, making it the initial imaging modality for patients presenting with suspected cardiovascular disease. Augmenting echocardiography with AI would not only allow for automated measurement of routine echocardiographic parameters such as the ejection fraction and improved daily efficiency,[7–9] but it might also facilitate identification of diseases not easily recognized by echocardiography such as cardiac amyloidosis.[10] Furthermore, as point-of-care ultrasound (POCUS) imaging is being increasingly utilized to evaluate

* Corresponding author. Tel: 50726848612, Fax: 5072667929, Email: attia.zachi@mayo.edu
†The first two authors contributed equally to the study.

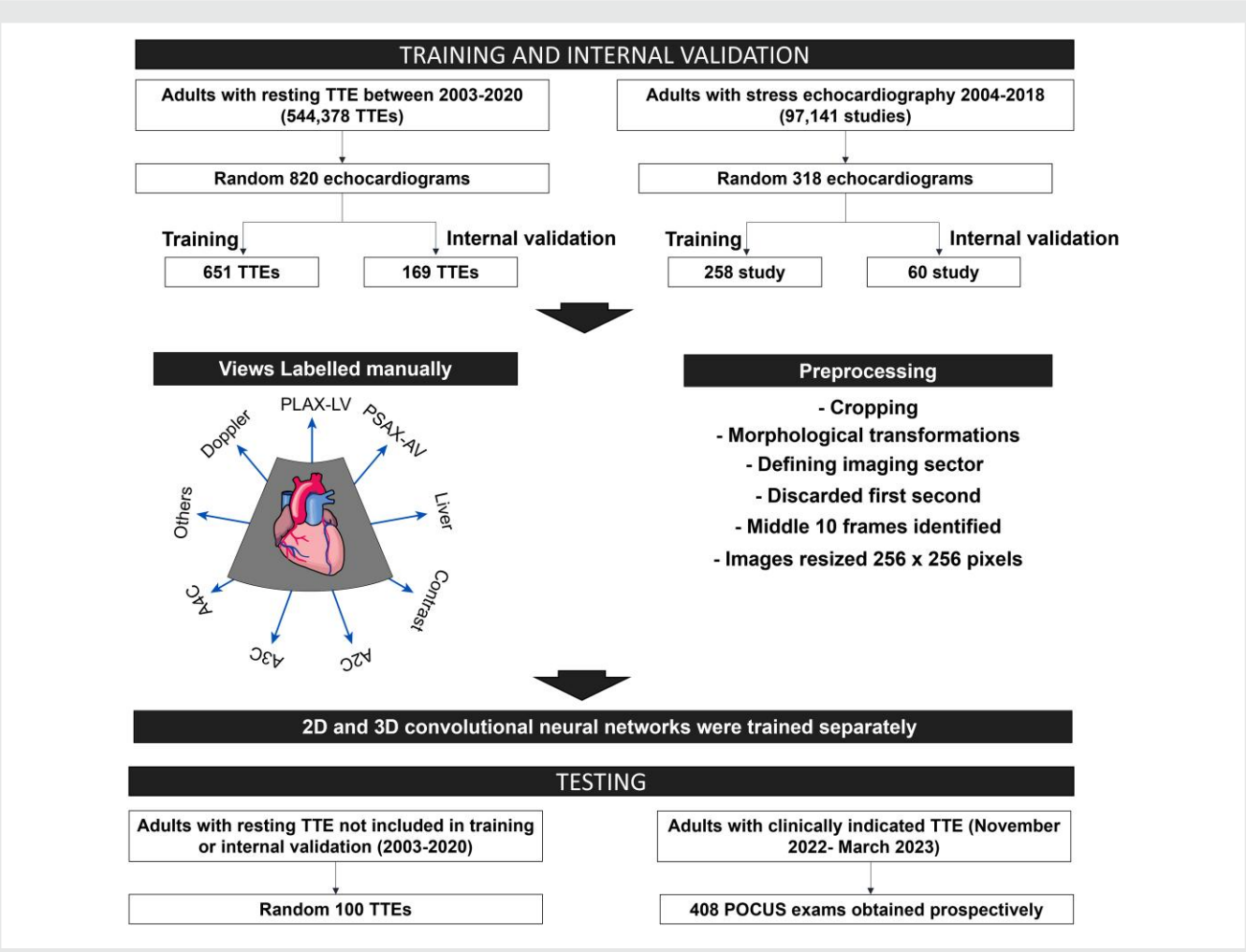


Figure 1 Overview of study design. 2D, two-dimensional; 3D, three-dimensional; A2C, apical two-chamber; A3C, apical three-chamber; A4C, apical four-chamber; PLAX-LV, parasternal long-axis left ventricle; POCUS, point-of-care ultrasound; PSAX-AV, parasternal short-axis aortic valve; TTE, transthoracic echocardiography.

cardiac structure and function in various clinical settings, ranging from primary care clinics to the intensive care units, enhancing POCUS with AI can improve the assessment accuracy of relevant cardiac parameters that can affect management strategies in these settings.

There are multiple challenges to the application of deep learning to the echocardiogram. First, a TTE study has different views, each providing different information; these views can be obtained in different order, and each view may be obtained multiple times depending on the sonographer and situation. This makes it different from a computed tomography or magnetic resonance imaging study, where the acquisition of videos is more standardized and the order and content of the videos incorporating each study can be predicted. Thus, deep learning cannot be applied immediately to a TTE study when specific views are required. Second, the quality of echocardiographic videos is dependent on the operator, the patient (e.g. body habitus, lung disease, and patient positioning), and the ultrasound probe and machine and can vary significantly between different TTE studies. This second challenge is even more pronounced in POCUS since images can be obtained in less optimal settings and by less experienced operators and the portable ultrasound machines are less sophisticated. While the latter challenge is mostly non-modifiable, the first challenge can be overcome by using an automated view classifier that can identify specific views of interest in each echocardiographic study.

There have been multiple automated TTE view classifiers developed previously.[11–15] A limitation in many of those view classifiers was that the testing sets contained pre-selected views of interest and did not necessarily cover all the potential views included in a routine TTE study. Therefore, the real-world performance of those view classifiers is unknown. Furthermore, none of the previous studies evaluated performance on echocardiographic views obtained using POCUS, which could have significant implications in clinical practice (e.g. early identification of cardiac disease and referral to cardiology in primary care clinics).

In this study, we set forth to (i) develop an automated convolutional neural network (CNN)-based view classifier that can identify major TTE views from real-life, full echocardiographic studies and (ii) evaluate the performance of the view classifier on cardiac videos obtained using POCUS.

Methods

Study population and design

Consent

The study was approved by the Institutional Review Board, which waived informed consent requirements. Only patients who had previously agreed

to allow access to their records for research were included. The data are available from the corresponding author upon a reasonable request.

Training and internal validation

The design of the study is illustrated in *Figure 1*. The TTEs of all adult patients who had a resting TTE study at Mayo Clinic, Rochester between 2003 and 2020 were identified (544 378 TTEs; the resting TTE cohort). Subsequently, a random sample of TTEs of ~45 unique patients per year was selected from this cohort for a total of 820 TTEs. Of those, 651 patients were used in training and 169 in internal validation in approximately a 4:1 ratio. To increase generalizability, all stress echocardiograms at Mayo Clinic, Rochester between 2004 and 2018 were also identified (97 141 studies; the stress echocardiography cohort). A simple random sample of 318 studies was then chosen. Of these, 258 TTEs were used in training and 60 in internal validation (approximately 4:1 ratio). Only one echocardiographic study for each patient was included in the combined cohort.

Testing

Testing using full transthoracic echocardiographic studies

A random sample of five to six TTEs per year (total of 100 patients) was chosen randomly from all adult patients with a TTE between 2003 and 2020 who were not selected for training or internal validation; no more than one TTE study was included for each patient; all videos obtained from the same patient were exclusively assigned to the training, validation, or testing groups.

Testing using point-of-care ultrasound

Cardiac videos were obtained prospectively by certified sonographers using POCUS from 408 random patients who had a clinical indication for a full outpatient TTE study between November 2022 and March 2023 (Lumify, Philips Healthcare). The videos were stored using the web-based platform Q-path, Telexy Healthcare, Port Coquitlam, Canada and were used for testing of the view classifier as a proof of concept to evaluate if the view classifier would be able to recognize and categorize views obtained using handheld ultrasound. The model was evaluated on the POCUS data set without any further training, which allowed us to test the robustness and generalizability of the model.

Testing using an independent cohort

We also evaluated the performance of the view classifier on a random 56 full TTE studies performed at Mayo Clinic in Arizona and Florida from a total of 460 000 TTEs between 2003 and 2021 as an independent sample. Only one TTE per patient was considered.

Echocardiography

Views were classified into the following nine categories: (i) parasternal long-axis left ventricle (PLAX-LV); (ii) parasternal short-axis aortic valve (PSAX-AV), which included both standard and zoomed-in views; (iii) apical four-chamber (A4C), which also included A4C right ventricle-focused view; (iv) apical three-chamber (A3C); (v) apical two-chamber (A2C); (vi) inferior vena cava/abdominal aorta, or liver, views; (vii) all views that include colour Doppler; (viii) all views that include contrast enhancement; and (ix) other remaining views, which included all other TTE views. Examples of the view categories are shown in Supplementary material online, *Figure S1*. An ultrasound image-enhancing agent (i.e. contrast) was used for endocardial border detection according to clinical practice recommendations, when two or more left ventricular segments could not be adequately visualized.

Transthoracic echocardiographic images from Mayo Clinic, Rochester were reviewed and labelled manually by one reviewer (J.A.N.), a cardiology fellow trained in view classification. Transthoracic echocardiographic images from the Mayo Clinic in Arizona and Florida were reviewed and labelled manually by a board-certified cardiologist (G.T.). All above views were identified and included from the resting TTE cohort resulting in a total number of 8862 videos in the training set and 2274 videos in the internal validation set (combined number of 11 136 videos), Supplementary material online, *Table S1*. The specific views included from the stress echocardiography cohort were PLAX-LV, A4C, A3C, A2C, and other remaining views [specifically including PSAX-mitral valve level (PSAX-MV), PSAX-papillary muscle level (PSAX-PAP), and PSAX-apex level] with 1407 videos used in training and 308 in internal validation, Supplementary material online, *Table S2*. In the training and internal validation sets, only one example of each view was included.

Transthoracic echocardiographic volumes with <10 frames including the still images of continuous wave Doppler, pulsed wave Doppler, tissue Doppler imaging, and M-mode were excluded in the data pre-processing stage (see next section). Other exclusion criteria included TTE studies acquired with a congenital protocol and cine videos with a view transforming to another view. The TTE images included both greyscale and B-mode colours (example in Supplementary material online, *Figure S1I*). Different clips for the same view may have different quality, depth, or gain (see Supplementary material online, *Figure S2*).

The views obtained using POCUS, which were used for testing the view classifier, were A2C, A3C, A4C, PLAX-LV, and PSAX-MV/PSAX-PAP (included in the 'others' category) views.

Data pre-processing

Institutional TTEs were stored in the Digital Imaging and COmmunication in Medicine (DICOM) format. The Echo Notion Software (Notion PACS, 2019) 'Batch Query' function was used to anonymize and download the TTEs in our study. Subsequently, the manually labelled views were identified using accession, series, and instance numbers, which collectively uniquely identified specific cine images.

The downloaded DICOM files were then cropped in an automated way to show the echocardiographic window with unnecessary information removed; this process resulted in the exclusion of company logos, heart rate, blood pressure, frame rate, and study date. Then, to isolate the imaging sector, the changing portion of the video across the temporal dimension (i.e. pixel values changing from frame to frame) was identified, and morphological transformations were applied. Finally, a bounding box was drawn around the largest moving portion of the video in order to define the imaging sector. Note that the ECG signal was removed if it was outside the sector but was retained if it overlapped the imaging sector in any way; this was done to avoid introducing any bias to resultant videos (see Supplementary material online, *Figure S3*). The initial second from each cine image was discarded as it was usually the least stable part of the study. Subsequently, 10 consecutive frames from the middle portion of each cine loop were used. All images were resized to 256 × 256 pixels with padding to maintain the aspect ratio. All pre-processing steps used Python 3.6 with pydicom 2.3, pillow 8.3, and opencv-python 4.5.

Convolutional neural networks

We first trained the view classifier using the two-dimensional (2D) ResNet-18 CNN.[16] For this purpose, the middle 10 frames of each cine loop were treated as still images and used as input data. Because the temporal dimension was not fed into the network in this training, the network was not aware of the consecutive nature of the frames and treated them as independent images. The 2D CNN model generated a score between 0 and 1 for each frame, for each of the nine classifications, with high values indicating high confidence that the image belongs in that category. The predicted category for each cine loop is the category with the highest average score. We also tested the 2D network using only one frame per cine loop to test its robustness.

We then trained another version of the view classifier using a three-dimensional (3D) ResNet-18 CNN to evaluate whether this would improve the performance of the view classifier. The 3D CNN took into account the third, temporal, dimension with Conv3d. This allows the 3D CNN model to incorporate the relationship between frames of the same cine loop. Similar to the 2D network, the 10 middle frames were used, but in this case, the 10 frames were treated as one example rather than 10 independent examples.

Additionally, for the 2D CNN, we implemented other widely used architectures, including Inception-v3,[16] VGG-13, and VGG-16.[17]

All 2D and 3D CNNs were trained with batch size 128, learning rate 0.01, stochastic gradient descent optimizer, and multi-step learning rate scheduler using Python 3.6 and PyTorch 1.10. PyTorch functions Conv2d and Conv3d were used for the convolution of 2D and 3D CNNs, respectively. Also, we used random resized cropping for data augmentation with a crop of random size ranging from 0.25 to 1.0 and a random aspect ratio ranging from 0.75 to 1.0. We trained the models for 200 epochs and selected the best model based on the validation performance with cross-entropy loss.

Model evaluation

Diagnostic performance was evaluated using accuracy, positive predictive value (PPV), and area under the receiver operating characteristic curve (AUC). When multiple examples existed for the same view in the testing set, only the two examples with the highest probability numbers were used. In the POCUS data set, there was only one example of each view obtained so the probability of that example was used. Overall accuracy was calculated as the number of correctly identified views overall divided by the total number of views. The per-view accuracy was calculated as the number of correctly identified videos from all available videos for each view. The PPV for each view was calculated as the number of correctly identified videos divided by the total number of predicted videos for each view. The PPV was considered to be the performance measure that relates the most to the intended use of the view classifier (i.e. to have correctly identified views of interest). The per-view and the averaged AUCs were reported. A confusion matrix was used to show correct and incorrect view classifications of the testing set.

The top two examples for each view in each TTE study identified by the 2D CNN were compared with the two top examples for each view that an expert echocardiographer (S.V.P.) identified manually. This comparison was performed in a random sample of patients who had four to five examples of a view of interest (including A2C, A3C, A4C, PLAX-LV, and PSAX-AV) resulting in a total of 73 overall videos (three different TTE studies for each of the mentioned five views were used).

Results

Overall, 909 patients were used in training, 229 in internal validation, 100 in testing using full TTE studies, and 408 in additional testing using POCUS. Patients had a wide range of ages, sizes, and comorbidities (*Table 1*; Supplementary material online, *Table S3*). Clips with varying depths and quality were used (Supplementary material online, *Figures S1 and S2*). Only seven TTE studies utilized contrast. The indications for the included TTE studies are shown in Supplementary material online, *Table S4*.

Performance of view classifier

The overall accuracy of the 2D CNN using the middle 10 consecutive frames was 97.3%, and the averaged AUC was 0.998 (*Figure 2*). The PPV values exceeded 93% for all views of interest (A2C, A3C, A4C, PLAX-LV, and PSAX-AV) (*Figure 2*). Description of the 40/1500 (2.7%) incorrectly identified views is shown in Supplementary material online, *Table S5*. A few observations could be made from the misclassified views. First, when 2D Doppler views were misidentified as another view, it was the correct 2D view to which the Doppler colour had been applied in 75% of cases. Second, contrast-enhanced videos were frequently identified as A2C, A3C, or A4C views. Third, there was confusion between the PSAX-AV view and other short-axis views (e.g. PSAX-MV) as well as other views that include the aortic valve (e.g. PLAX-zAV and ascending aorta). Finally, there were some rare occurrences in the test set that the view classifier was not trained to classify (e.g. pulmonary artery main branches, prosthetic mitral valve with shadowing, severely enlarged right ventricle, among others).

Because each frame in a cine loop was treated independently, some frames belonging to the same view were identified as two or more different views by the 2D CNN in the testing set. However, the use of the averaged probability from all included frames for each view and then the use of the two examples with the highest probability for each view were strategies utilized to help reduce the impact of this issue. This was reflected in a lower accuracy of the 2D CNN when testing using only one random frame (accuracy 95.8%; Supplementary material online, *Figure S4*).

The average AUCs for 2D CNNs with the Inception-v3, VGG-13, and VGG-16 architectures were 0.998, 0.997, and 0.997, respectively, and they were almost equivalent to the model using the ResNet-18 architecture. Therefore, we used ResNet-18 architecture for the 3D CNN because of the equivalent performance and the computationally lighter architecture in terms of the number of trainable parameters.

On the other hand, the overall accuracy of the 3D CNN was 96.3% with an averaged AUC of 0.998 (*Figure 3*). Similar to the 2D CNN, the per-view accuracy exceeded 98% in all views. However, the PPV values were lower compared with the 2D CNN although they still exceeded 89% for these views (A2C, A3C, A4C, PLAX-LV, and PSAX-AV) (*Figure 3*). Description of the 56/1495 (3.7%) incorrectly identified views is shown in Supplementary material online, *Table S6*. There were more incorrectly predicted PSAX-AV views compared with the 2D CNN (20 vs. 11 videos, respectively), but, similar to the 2D CNN, the confusion was often with views that contained the ascending aorta, PLAX-zAV view, or other short-axis views. In other instances, the PLAX-RV inflow and the Sc4C-septum views were also misidentified as PSAX-AV. Second, contrast-enhanced videos were also frequently identified as A2C, A3C, or A4C views. Third, there were several 2D Doppler mode videos misidentified as other views, although in six of nine cases, these were the views of the underlying 2D videos to which the Doppler colour was applied.

Notably, almost all of the misclassified views by the 2D CNN had a probability number/inference score of <90% (Supplementary material online, *Figure S5*). This was not necessarily the case in the 3D CNN, where many more incorrectly labelled videos had probabilities $\geq$ 90%.

Interestingly, when compared with the top two views that an expert echocardiographer identified in a random sample with four to five examples of a view of interest (A2C, A3C, A4C, PLAX-LV, and PSAX-AV), there was agreement in 29 of 30 videos selected as top two examples (97%). Specifically, all top two examples per view were the same for the three TTEs in each of the A3C, A4C, PLAX-LV, and PSAX AV views. In the A2C view, there was disagreement in the second top example in one of the three TTE studies (one of six examples). The agreement with the single best example was 100%.

Saliency maps

To be able to understand the basis on which the CNNs classified the views to different categories, we looked at saliency mapping, which represents the input pixels that the CNNs found to be shared by examples of the same view and used as identifying features of that certain view. Since the 2D CNN performed better than the 3D CNN, we show representative examples of saliency mapping across different views as identified by 2D CNNs in Supplementary material online, *Figure S6*. These features were consistent across different images from the same views and are similar to those used by echocardiographers to classify the views.

Prospective testing using point-of-care ultrasound videos

In the POCUS data set, there were 32 of 2039 views were misidentified using the 2D CNN with an overall accuracy of 98.4% and an averaged AUC of 0.998 (*Figure 4*). The PPV exceeded 95% in all the five view categories. On the other hand, there were 102/2039 misidentified views when using the 3D CNN with an overall accuracy of 95.0% and an averaged AUC of 0.996 (*Figure 5*). Positive predictive value exceeded 91% in all view categories with the least value being for the A4C view where 35 (8%) of the images of the view classifier identified belonged to the A4C and 3 (0.7%) to the A2C view. Notably, PPV was universally better using the 2D CNN for all views tested (A2C, A3C, A4C, PLAX-LV, and others) (*Figures 4* and *5*).

Performance in the independent cohort

In random 56 full TTE studies from the Mayo Clinic in Arizona and Florida, the 2D CNN still performed well with an overall accuracy of 96.62% (Supplementary material online, *Figure S7*).

Table 1 Characteristics of patients in the training, internal validation, and transthoracic echocardiographic testing sets

	Training ($n = 909$)	Validation ($n = 229$)	Testing ($n = 100$)
Sex (male)	487 (53.6%)	109 (47.6%)	46 (46%)
Age at echocardiogram	69 (57, 78)	68 (59, 77)	65 (50, 73)
Ethnicity			
Hispanic or Latino	8 (0.9%)	4 (1.7%)	1 (1%)
Non-Hispanic or Latino	732 (80.5%)	190 (83%)	76 (76%)
N-Miss	169	35	23
Race			
American Indian/Alaskan Native	2 (0.2%)	2 (0.9%)	0 (0%)
Asian	11 (1.2%)	4 (1.7%)	1 (1%)
Black or African American	11 (1.2%)	4 (1.7%)	4 (4%)
Native Hawaiian/Pacific Islander	1 (0.1%)	0 (0%)	0 (0%)
White	823 (90.5%)	207 (90.4%)	89 (89%)
Other	13 (1.4%)	3 (1.3%)	1 (1%)
N-Miss	48	9	5
Year			
2003–05	186	39	16
2006–08	173	47	16
2009–11	146	43	22
2012–14	151	32	15
2015–17	131	48	15
2018–20	122	20	17
Vendors	251	61	13
Acuson	329	81	44
GE	265	73	28
Philips	1	—	1
Toshiba	63	—	14
Unknown			
Weight	85.25 (71.84, 108.22)	84.0 (68.0, 105.6)	97.0 (77.87, 127.5)
Height	168.0 (158, 176.65)	166.5 (159, 176.25)	169.9 (159.45, 177.8)
BMI (kg/m^2)	29.1 (25.01, 36.42)	29.76 (25, 35.79)	31.96 (26.82, 45.59)
Congestive heart failure	150 (16.5%)	36 (15.9%)	21 (21%)
Peripheral vascular disorders	166 (18.3%)	47 (20.7%)	16 (16%)
Chronic obstructive pulmonary disease	131 (14.4%)	37 (16.3%)	26 (26%)
Diabetes mellitus	139 (15.3%)	38 (16.7%)	15 (15%)
Renal failure	81 (8.9%)	21 (9.3%)	9 (9%)
Liver disease	41 (4.5%)	14 (6.2%)	7 (7%)
Atrial fibrillation	186 (20.5%)	50 (22%)	20 (20%)
Hyperlipidaemia	340 (37.5%)	90 (39.6%)	35 (35%)
Hypertension	417 (46%)	104 (45.8%)	43 (43%)
Pulmonary hypertension	71 (7.8%)	16 (7.0%)	17 (17%)

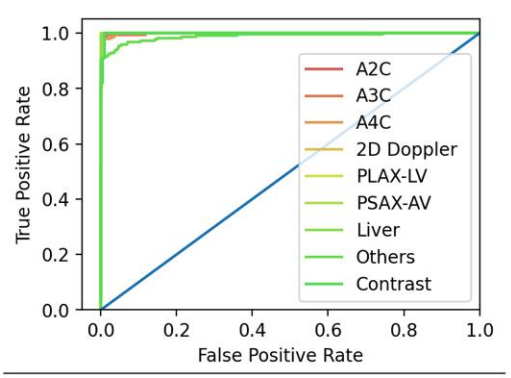
Continuous variables are summarized as median (interquartile range). Categorical variables are summarized as count (%).

Performance according to decade of study

We assessed the performance of the 2D CNN view classifier on TTEs from the testing set at Mayo Clinic, Rochester obtained at/before vs. after 10 October 2012 (Supplementary material online, *Figure S8*). Performance was largely similar with accuracy of 96.7% and 98.1%, respectively.

Inter-observer variability in labelling transthoracic echocardiographic studies from Mayo Clinic, Rochester

A total of random 45 examples were validated by an experienced level III board-certified echocardiologist (S.V.P.) with 100% agreement with the labelling performed by (J.A.N.).



View	AUROC	Acc	PPV	NPV	Sen	Spe
A2C	1.00	0.99	0.94	1.00	0.98	0.99
A3C	1.00	0.99	0.96	1.00	0.97	0.99
A4C	1.00	1.00	0.98	1.00	0.98	1.00
2D Doppler	1.00	1.00	1.00	1.00	0.98	1.00
PLAX-LV	1.00	1.00	1.00	1.00	1.00	1.00
PSAX-AV	1.00	0.99	0.93	1.00	1.00	0.99
Liver	1.00	1.00	0.99	1.00	1.00	1.00
Others	0.99	0.99	1.00	0.98	0.90	1.00
Contrast	1.00	1.00	0.64	1.00	0.82	1.00

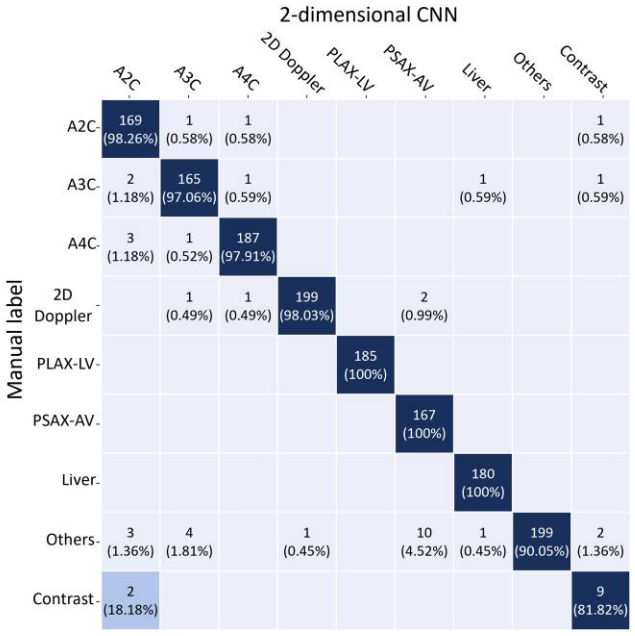


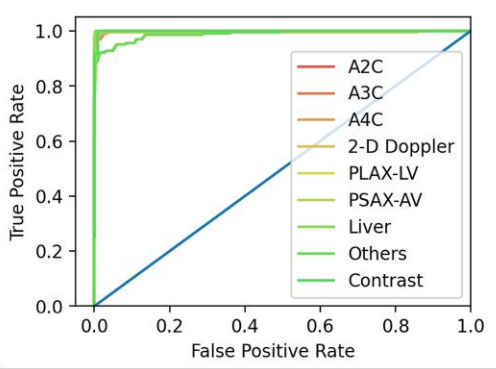
Figure 2 Performance of the two-dimensional convolutional neural network on the transthoracic echocardiographic testing data set. Left graph shows the receiver operating characteristic curves for each of the nine view categories and the area under receiver operating characteristic curve, accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of each of the view categories. Right graph shows the confusion matrix for the model. 2D, two-dimensional; A2C, apical two-chamber; A3C, apical three-chamber; A4C, apical four-chamber; Acc, accuracy; CNN, convolutional neural network; NPV, negative predictive value; PLAX-LV, parasternal long-axis left ventricle; POCUS, point-of-care ultrasound; PPV, positive predictive value; PSAX-AV, parasternal short-axis aortic valve; Sen, sensitivity; Spe, specificity; TTE, transthoracic echocardiography.

Discussion

An automated view classifier that can identify specific views of interest from full TTE and POCUS studies provides the building blocks that will allow for the application of deep neural networks to echocardiography exams. This is especially the case since the performance of deep neural networks depends largely on both the quality and the number of input exams used for training, the latter of which would otherwise be a cumbersome task to perform at a large scale manually. In this study, we developed an automated CNN-based TTE view classifier that identified five major echocardiographic views (i.e. PLAX-LV, PSAX-AV, A4C, A3C, and A2C) with an excellent accuracy exceeding 97% and an AUC exceeding 0.998. The accuracy was still high at 96.6% in an independent cohort from Mayo Clinic in Arizona and Florida. Other views including the subcostal views, contrast-enhanced views, 2D Doppler views, and others had to be labelled to allow for testing on a complete full TTE study, simulating a real-world experience. Importantly, the developed view classifier performed comparably when applied to videos obtained using a handheld ultrasound device (POCUS).

We utilized 2D and 3D CNNs separately for the classification of echocardiographic views. In general, the 2D network offered the advantage of being lighter and faster to execute when compared with the 3D network. Additionally, it was more versatile as it could accommodate any number of frames and could operate effectively even when presented using a single frame with an accuracy of 95.8%. However, it did not account for the temporal dimension of cine loops. Therefore, it analysed the different frames from the same cine loop as independent data points. One consequence we observed was that two different frames from the same cine loop in the testing set could be identified as two different views. We were able to overcome such a consequence by averaging the probability output number from the middle 10 frames in each cine loop, which allowed only one view to be predicted from each loop. This resulted in enhanced performance when compared with using only one frame to predict the view of interest. Following this step by choosing the two videos with the highest probability number/inference score for each view in each study helped further mitigate this challenge. Interestingly and despite all these sensible limitations for a 2D CNN, the performance of the view classifier, as indicated by the PPV, was similar or better in the identification of the A2C, A3C, PSAX-AV, and PLAX-LV views when using the 2D CNN compared with the 3D CNN. Although the 3D CNN seemed to have a superior performance in the identification of the A4C view while using the TTE testing set, the 2D CNN still performed better in the identification of the A4C view in the POCUS testing set. This could be because of the increased input used for training the 2D in our study (10 frames from each video vs. one video for the 3D CNN). Two-dimensional CNNs were also found to outperform 3D CNNs in view classification of echocardiographic videos in a previous study.[16] Reassuringly, the 2D CNN was found to use features similar to those used by echocardiographers to classify the views, as shown using saliency mapping.

Different architectures have been used in other recently developed echocardiographic view classifiers. For example, Gao *et al.*[14] used two 2D CNN networks, one of which was concentrated along spatial dimensions, and the other along the temporal dimension. Each network was executed separately, but input from both the spatial and temporal networks was combined to obtain final classification scores for eight views of interest. The classifier performed with an overall accuracy of 92.1% in distinguishing the eight views from each other. Kusunose *et al.*[13] instead used one 2D CNN, similar to ours, with different types of input such as



View	AUROC	Acc	PPV	NPV	Sen	Spe
A2C	1.00	0.99	0.90	1.00	0.99	0.99
A3C	1.00	0.99	0.93	0.99	0.96	0.99
A4C	1.00	1.00	0.99	1.00	0.97	1.00
2-D Doppler	1.00	0.99	1.00	0.99	0.96	1.00
PLAX-LV	1.00	1.00	0.99	1.00	0.99	1.00
PSAX-AV	1.00	0.99	0.89	1.00	0.99	0.98
Liver	1.00	1.00	0.98	1.00	1.00	1.00
Others	0.99	0.98	0.99	0.98	0.88	1.00
Contrast	1.00	1.00	0.91	1.00	0.83	1.00

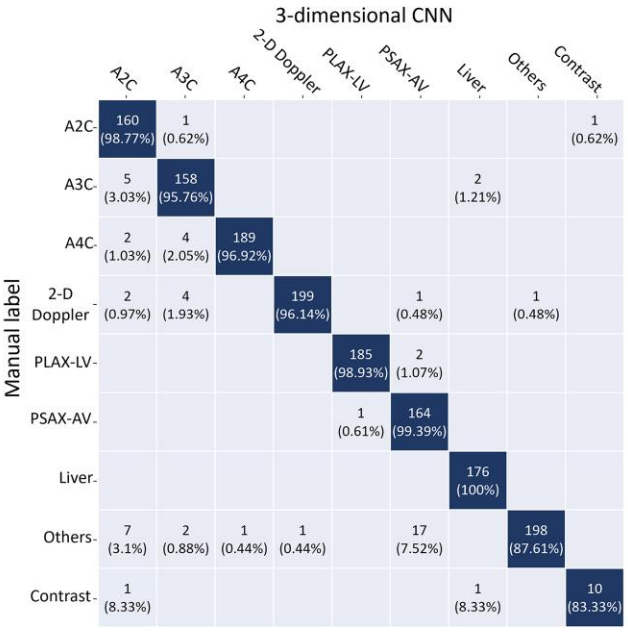
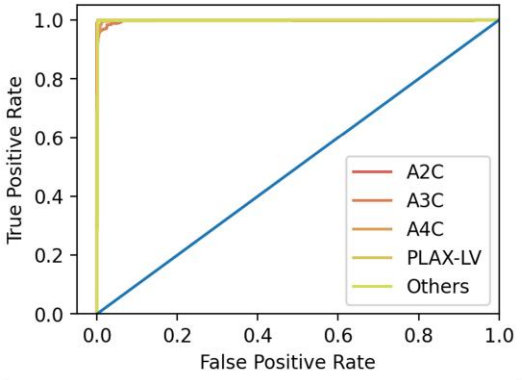


Figure 3 Performance of the three-dimensional convolutional neural network on the transthoracic echocardiographic testing data set. Left graph shows the receiver operating characteristic curves for each of the nine view categories and the area under receiver operating characteristic curve, accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of each of the view categories. Right graph shows the confusion matrix for the model. 2D, two-dimensional; A2C, apical two-chamber; A3C, apical three-chamber; A4C, apical four-chamber; Acc, accuracy; CNN, convolutional neural network; NPV, negative predictive value; PLAX-LV, parasternal long-axis left ventricle; POCUS, point-of-care ultrasound; PPV, positive predictive value; PSAX-AV, parasternal short-axis aortic valve; Sen, sensitivity; Spe, specificity; TTE, transthoracic echocardiography.



View	AUROC	Acc	PPV	NPV	Sen	Spe
A2C	1.00	1.00	1.00	1.00	1.00	1.00
A3C	1.00	0.99	0.99	0.98	0.94	1.00
A4C	1.00	0.99	0.95	1.00	0.99	0.99
PLAX-LV	1.00	1.00	1.00	1.00	1.00	1.00
Others	1.00	1.00	0.99	1.00	1.00	0.99

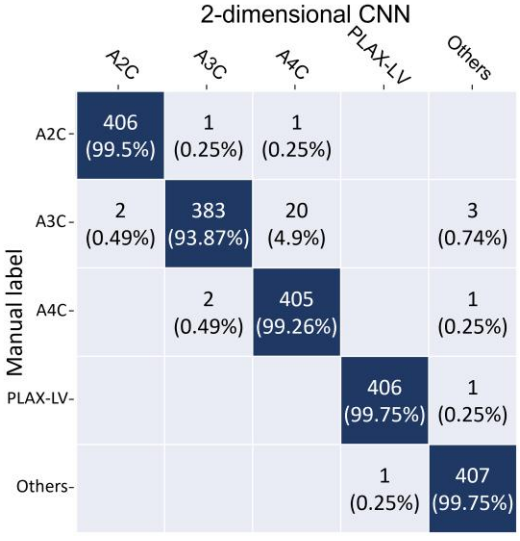
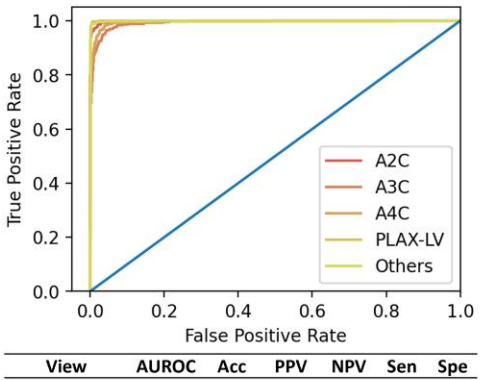


Figure 4 Performance of the two-dimensional convolutional neural network on the point-of-care ultrasound data set. Left graph shows the receiver operating characteristic curves for each of the nine view categories and the area under receiver operating characteristic curve, accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of each of the view categories. Right graph shows the confusion matrix for the model. A2C, apical two-chamber; A3C, apical three-chamber; A4C, apical four-chamber; Acc, accuracy; CNN, convolutional neural network; NPV, negative predictive value; PLAX-LV, parasternal long-axis left ventricle; POCUS, point-of-care ultrasound; PPV, positive predictive value; PSAX-AV, parasternal short-axis aortic valve; Sen, sensitivity; Spe, specificity; TTE, transthoracic echocardiography.



View	AUROC	Acc	PPV	NPV	Sen	Spe
A2C	1.00	0.99	0.97	0.99	0.96	0.99
A3C	0.99	0.97	0.96	0.97	0.87	0.99
A4C	0.99	0.97	0.91	0.99	0.95	0.98
PLAX-LV	1.00	0.99	0.99	0.99	0.98	1.00
Others	1.00	0.98	0.94	1.00	0.99	0.98

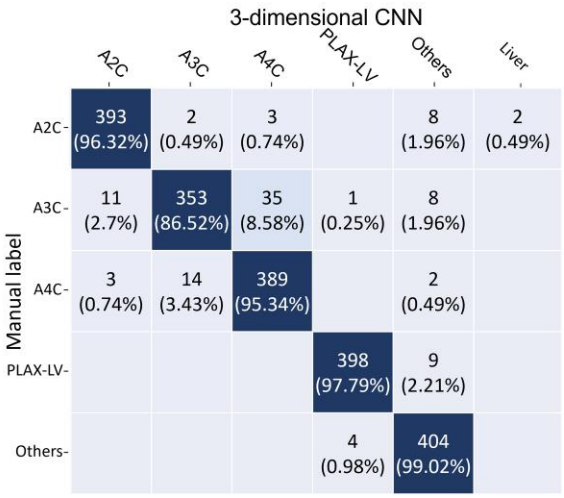


Figure 5 Performance of the three-dimensional receiver operating characteristic on the point-of-care ultrasound data set. Left graph shows the receiver operating characteristic curves for each of the nine view categories and the area under receiver operating characteristic curve, accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of each of the view categories. Right graph shows the confusion matrix for the model. A2C, apical two-chamber; A3C, apical three-chamber; A4C, apical four-chamber; Acc, accuracy; CNN, convolutional neural network; NPV, negative predictive value; PLAX-LV, parasternal long-axis left ventricle; POCUS, point-of-care ultrasound; PPV, positive predictive value; PSAX-AV, parasternal short-axis aortic valve; Sen, sensitivity; Spe, specificity; TTE, transthoracic echocardiography.

images averaged over time and 10 uniformly spaced images in one cardiac cycle according to a semi-automatic heartbeat algorithm. The performance of the network was the best when utilizing the 10 selected images with an overall accuracy of 98.1% in the classification of five views of interest. Khamis *et al.*[15] used a sequential classification method that combined a spatiotemporal feature extraction method and a dictionary learning method to predict three views with an overall accuracy of 95%. Zhang *et al.*[14] trained a Visual Geometry Group (VGG) network to predict 23 view classes. The final output score was averaged for 10 randomly selected frames from each video. The performance was lower compared with other view classifiers that might be related to the comprehensive number of target view classes. Howard *et al.*[18] developed a deep learning model combining two CNNs taking spatial and temporal stream, respectively, with video data set. The model showed a lower error rate compared with single 2D or 3D CNN, and the improved error rate was comparable with our performance. Madani *et al.*[12] suggested a 2D CNN inspired by VGG-16, and it also demonstrated that a simple majority vote for video classification with a 2D model could be effective. The performance was consistent with ours in 27 test echocardiographic studies. Our study utilized 2D and 3D CNNs without the use of additional methods to extract features along the temporal axis and was able to classify nine view categories with an overall accuracy of 96%.

Since each study used different patients, number of views, and preprocessing methods, direct comparison between the networks is not feasible. However, there are many important characteristics for this current TTE view classifier. First, we included TTE studies across a wide time interval spanning 18 years (2003–2020) and representing a variable patient population with different ages, sizes, comorbidities, and a wide range of ultrasound vendors. This allowed the view classifier to be exposed to many variations of the same view, different patient characteristics, and variable image quality. Second, we assessed the performance of the view classifier on real-world full TTE studies. This is different from multiple previous TTE view classifiers that were tested using sets that only contained pre-selected views of interest,[13–15] in which the performance could change drastically when exposed to 'new', unfamiliar views in full TTE studies. Importantly, we were also able to assess the performance of the current view classifier on a number of echocardiographic views obtained using POCUS with excellent accuracy and PPVs. Notably, no previous study has evaluated the performance of a view classifier on POCUS images.

Our focus was to correctly identify five views of interest (PLAX-LV, PSAX-AV, A4C, A3C, and A2C), as these views would allow for the characterization of the function and disease processes involving all major cardiac structures including (i) the left atrium, left ventricle, and mitral valve (PLAX-LV, A4C, A3C, and A2C); (ii) right atrium, right ventricle, and tricuspid valve (A4C); and (iii) aortic valve (PSAX-AV). With that said, the view classifier can be further tailored to identify additional views pertinent for certain disease processes (e.g. PLAX-zoomed AV view in studying aortic stenosis). Additionally, we chose to separate views containing Doppler colour or those enhanced by contrast from their corresponding views to avoid distracting features when training future CNNs to identify shared disease patterns. Despite this, on occasion, Doppler colour was missed by the view classifier, and views with Doppler colour were still identified as their baseline corresponding views. Notably, the PLAX-LV and A3C views have many similarities; they include the anteroseptal and inferolateral LV walls, LV outflow tract, the aortic valve, the mitral valve, and the RV outflow tract. However, the A3C view shows the LV outflow tract in a more parallel fashion to the insonation angle that would allow for studying flow in the LV outflow tract. It is possible that the CNN is able to appreciate these differences and, therefore, overlap between these two views was not an issue in this current view classifier. Reassuringly, when compared with the top two views that an expert echocardiographer identified, there was agreement in 29 of 30 videos selected as top two examples (97%) by the view classifier.

The current view classifier is expected to pave the way for the application of CNNs to the echocardiogram not only for automated assessment of routinely obtained echocardiographic parameters but also to identify disease patterns not easily recognized by the human eye using routine echocardiography, such as cardiac amyloidosis,

sarcoidosis, storage diseases, and subclinical forms of hypertrophic cardiomyopathies. The model demonstrated exceptional accuracy on the POCUS data set, exhibiting robustness without the need for re-training or fine-tuning. This highlights the potential for AI-based algorithms to be widely applied to this tool, with limitless future applications anticipated, especially as the use of POCUS is exponentially increasing and becoming an integral part of the 'physical exam'.[19,20] For example, the future development of automated neural networks to identify low left ventricular ejection fraction or impaired right ventricular systolic dysfunction on a quick POCUS exam would provide valuable information to clinicians not certified in echocardiography. This can facilitate referral to full TTE studies and cardiology consultation in primary care clinics and would allow for earlier initiation of available therapies with proven effectiveness in heart failure. Similarly, the assessment of left and right ventricular function in the intensive care units can have an immediate impact on management strategies minute to minute while awaiting a formal TTE study. Other potential applications of the AI-enhanced POCUS include screening for diseases in the community and in rural settings where the accessibility to a full TTE exam might be limited, such as screening for severe aortic stenosis and heart failure in patients with dyspnoea. In this scenario, the AI-enhanced POCUS might help triage these patients and identify those in need to be transported to appropriate medical centres for further evaluation and management.

Limitations

All TTE studies were obtained from Mayo Clinic sites, and performance on echocardiographic studies obtained in other institutions needs further evaluation. Specifically, the Mayo Clinic format of the A4C view is left-right flipped compared with most other institutions (i.e. the left ventricle is on the left side of the image), and testing the ability of the view classifier to categorize the apical views in the standard format is warranted. However, it should be noted that the view classifier had excellent performance in geographically distinct locations of the Mayo Clinic in Arizona and Florida, even when the training only involved TTEs from the Mayo Clinic site in Rochester, Minnesota. Furthermore, the view classifier was validated using POCUS images where different ultrasound machines were utilized and maintained excellent performance without additional training on POCUS images. The inclusion of videos obtained by different vendors across a long time period spanning 20 years and testing using TTEs from other locations and POCUS proves good generalizability.

Because we utilized the middle 10 frames in each loop, we likely included frames in different stages of the cardiac cycle for each view due to the variability in the length and number of cardiac cycles included in each loop. Also, the cine rate, defined as the number of frames per second, of each TTE was not equivalent across all studies. However, this should provide a means for the CNNs to be exposed to more variations of the same view and might have helped in the identification of the most important features to classify a view (e.g. it should not matter whether the mitral valve is open or closed to identify the A4C view). Furthermore, we chose to avoid the first 10 frames whenever feasible since the initial part of the acquired cine loops is usually the least stable, although that would have provided a relatively consistent stage of the cardiac cycle.

Despite the high PPVs of the current view classifier, there will be a small portion of misidentified videos (e.g. 56 loops in each 1000 TTE studies identified for A2C) that may need manual review upon implementation of the classifier. However, this remains more manageable than having to review >70 videos for each TTE study (e.g. >70 000 videos for 1000 TTE studies to identify A2C views). Moreover, choosing views with certain probability numbers/inference scores (e.g. >0.90) might allow for full automatization of the view classification process. The testing TTE set had only seven studies utilizing contrast. A formal comparison of the view classifier performance between these seven studies and other higher-quality TTE studies was therefore not feasible. The testing set obtained using POCUS had only five view categories, and it is possible that this contributed to the higher accuracy of the model observed in these patients. Future validation of the view classifier on POCUS images obtained in less controlled environments and by less experienced operators is needed. Finally, given that contrast-enhanced loops are not routinely obtained in a TTE study (i.e. only obtained when two or more left ventricular segments could not be adequately visualized), we only had a small number of examples of contrast-enhanced loops in the training and validation sets, which likely affected the performance of the model for this view category.

Conclusions

An automated AI-based view classifier utilizing CNNs was able to classify cardiac views obtained using TTE from real-life, full echocardiographic studies with high accuracy exceeding 96% and averaged AUC exceeding 0.997. The performance on POCUS was comparable. Validation in other independent Mayo Clinic sites yielded similar results. This view classifier will allow for the future application of deep learning neural networks to cardiac images obtained by either TTE or POCUS to evaluate cardiac structure and function and to detect various disease processes. Future studies are needed to evaluate the performance of the view classifier on echocardiographic images obtained in other centres is needed.

Lead author biography

Jwan A. Naser, MBBS, is a Cardiology fellow in the Clinician Investigator at Mayo Clinic, Rochester. She finished her medical school at Jordan University of Science and Technology and her Internal Medicine residency at Mayo Clinic, Rochester. Her research interests include echocardiography, valvular heart diseases, atrial fibrillation, diastolic dysfunction, and artificial intelligence.

Eunjung Lee, PhD, is a senior data analyst at Mayo Clinic, Rochester, USA. She received her PhD degree in the Department of Intelligence and Information from Seoul National University, Seoul, South Korea, in 2021. Her general research interests are in the fields of data science and machine learning in cardiovascular medicine.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

None declared.

Conflict of interest: none declared.

Data availability

The data underlying the manuscript will be shared upon reasonable request to the corresponding author.

References

1. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**:2402–2410.
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;**542**:115–118.
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;**18**:500–510.
4. Soffer S, Klang E, Shimon O, Barash Y, Cahan N, Greenspana H, *et al.* Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Sci Rep* 2021;**11**:15814.
5. Jacobs C, Setio AAA, Scholten ET, Gerke PK, Bhattacharya HM, Hoesein FA, *et al.* Deep learning for lung cancer detection on screening CT scans: results of a large-scale public competition and an observer study with 11 radiologists. *Radiol Artif Intell* 2021;**3**:e210027.
6. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, *et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;**394**:861–867.
7. Knackstedt C, Bekkers SCAM, Schummers G, Schreckenberg M, Muraru D, Badano LP, *et al.* Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. *J Am Coll Cardiol* 2015;**66**:1456–1466.
8. Pellikka PA, Strom JB, Pajares-Hurtado GM, Keane MG, Khazan B, Qamruddin S, *et al.* Automated analysis of limited echocardiograms: feasibility and relationship to outcomes in COVID-19. *Front Cardiovasc Med* 2022;**9**:937068.
9. Salte IM, Østvik A, Smistad E, Melichova D, Nguyen TM, Karlsen S, *et al.* Artificial intelligence for automatic measurement of left ventricular strain in echocardiography. *JACC Cardiovasc Imaging* 2021;**14**:1918–1928.
10. Goto S, Mahara K, Beussink-Nelson L, Ikura H, Katsumata Y, Endo J, *et al.* Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nat Commun* 2021;**12**:2726.
11. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, *et al.* Fully automated echocardiogram interpretation in clinical practice. *Circulation* 2018;**138**:1623–1635.
12. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 2018;**1**:6.
13. Kusunose K, Haga A, Inoue M, Fukuda D, Yamada H, Sata M. Clinically feasible and accurate view classification of echocardiographic images using deep learning. *Biomolecules* 2020;**10**:665.
14. Gao Y, Zhu Y, Liu B, Hu Y, Yu G, Guo Y. Automated recognition of ultrasound cardiac views based on deep learning with graph constraint. *Diagnostics (Basel)* 2021;**11**:1177.
15. Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D. Automatic apical view classification of echocardiograms using a discriminative learning dictionary. *Med Image Anal* 2017;**36**:15–21.
16. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.p. 770–778.
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
18. Howard JP, Tan J, Shun-Shin MJ, Mahdi D, Nowbar AN, Arnold AD, *et al.* Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography. *J Med Artif Intell* 2020;**3**:4.
19. Lee L, DeCara JM. Point-of-care ultrasound. *Curr Cardiol Rep* 2020;**22**:149.
20. Mehta M, Jacobson T, Peters D, Le E, Chadderdon S, Allen AJ, *et al.* Handheld ultrasound vs. physical examination in patients referred for transthoracic echocardiography for a suspected cardiac condition. *JACC Cardiovasc Imaging*. 2014; **7**:983–990. doi: 10.1016/j.jcmg.2014.05.011.