Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research Paper

Spatio-temporal hybrid neural networks reduce erroneous human "judgement calls" in the diagnosis of Takotsubo syndrome

Fahim Zaman[a,#], Rakesh Ponnapureddy[b,#], Yi Grace Wang[c], Amanda Chang[b], Linda M Cadaret[b], Ahmed Abdelhamid[b], Shubha D Roy[b], Majesh Makan[d], Ruihai Zhou[e], Manju B Jayanna[f], Eric Gnall[f], Xuming Dai[g], Avneet Singh[h], Jingsheng Zheng[i], Venkata S Boppana[j], Feng Wang[k], Pahul Singh[l], Xiaodong Wu[a,*], Kan Liu[b,*]

[a] Department of Electrical and Electronic Engineering, University of Iowa, Iowa city, IA, United States
[b] Division of Cardiology, Department of Medicine, University of Iowa, Iowa City, IA, United States
[c] Department of Mathematics, California State University Dominguez Hills, Carson, CA, United States
[d] Division of Cardiology, Department of Medicine, Washington University in St. Louis, St. Louis, MO, United States
[e] Division of Cardiology, Department of Medicine, University of North Carolina, Chapel Hill, United States
[f] Division of Cardiology, Department of Medicine, Lankenau Medical Center, Wynnewood, PA, United States
[g] Department of Cardiology, New York Presbyterian Queens/Weill Cornell Medical College, New York City, NY, United States
[h] Division of Cardiology, Department of Medicine, State University of New York, Syracuse, NY, United States
[i] Department of Cardiology, AtlaniCare Regional Medical Center, Pomona, NJ, United States
[j] Division of Cardiology, Department of Medicine, University of Kansas-Wichita, Wichita, KS, United States
[k] Department of Cardiology, Providence Regional Medical Center, Washington State University, Everett, WA, United States
[l] Department of Cardiology, Northwest Health Medical Center, Bentonville, AR, United States

ARTICLE INFO

ABSTRACT

Background: We investigate whether deep learning (DL) neural networks can reduce erroneous human "judgment calls" on bedside echocardiograms and help distinguish Takotsubo syndrome (TTS) from anterior wall ST segment elevation myocardial infarction (STEMI).

Methods: We developed a single-channel (DCNN[2D SCI]), a multi-channel (DCNN[2D MCI]), and a 3-dimensional (DCNN[2D+t]) deep convolution neural network, and a recurrent neural network (RNN) based on 17,280 still-frame images and 540 videos from 2-dimensional echocardiograms in 10 years (1 January 2008 to 1 January 2018) retrospective cohort in University of Iowa (UI) and eight other medical centers. Echocardiograms from 450 UI patients were randomly divided into training and testing sets for internal training, testing, and model construction. Echocardiograms of 90 patients from the other medical centers were used for external validation to evaluate the model generalizability. A total of 49 board-certified human readers performed human-side classification on the same echocardiography dataset to compare the diagnostic performance and help data visualization.

Findings: The DCNN (2D SCI), DCNN (2D MCI), DCNN(2D+t), and RNN models established based on UI dataset for TTS versus STEMI prediction showed mean diagnostic accuracy 73%, 75%, 80%, and 75% respectively, and mean diagnostic accuracy of 74%, 74%, 77%, and 73%, respectively, on the external validation. DCNN(2D+t) (area under the curve [AUC] 0·787 vs. 0·699, $P = 0.015$) and RNN models (AUC 0·774 vs. 0·699, $P = 0.033$) outperformed human readers in differentiating TTS and STEMI by reducing human erroneous judgement calls on TTS.

Interpretation: Spatio-temporal hybrid DL neural networks reduce erroneous human "judgement calls" in distinguishing TTS from anterior wall STEMI based on bedside echocardiographic videos.

Funding: University of Iowa Obermann Center for Advanced Studies Interdisciplinary Research Grant, and Institute for Clinical and Translational Science Grant. National Institutes of Health Award (1R01EB025018−01).

* Corresponding authors.
   E-mail addresses: xiaodong-wu@uiowa.edu (X. Wu), kan-liu@uiowa.edu (K. Liu).
# F Zaman and R Ponnapureddy equally contributed to this study

Research in context

Evidence before this study

Echocardiography plays a vital role in triage and management of cardiovascular emergencies. A PubMed search for all types of papers in all languages up to May 28, 2021 with search terms of "echocardiography"(All Fields) AND "diagnosis"(All Fields) AND "deep learning"(All Fields) yielded 37 results, which have been focused on the investigation of cardiac pathology to help differential diagnosis of chronic cardiovascular disorders. The literature is lacking with regard to the studies that apply deep learning (DL) to real-time imaging for diagnosis or triage of acute cardiovascular disorders. Meanwhile, most of the reported DL prediction models were developed based on still-frame echocardiographic images with increased data yield and improved classifications, but showed variable performance in finding advanced diagnostic markers.

Added value of this study

We show that spatio-temporal hybrid DL neural networks reduce erroneous human "judgement calls" in distinguishing Takotsubo syndrome (TTS) from anterior wall ST segment elevation myocardial infarction based on bedside echocardiograms. Effective spatio-temporal modeling in real-time imaging can help triage cardiovascular emergencies and resolve time-sensitive diagnostic dilemmas. Our study also demonstrates the potential of DL neural networks to reduce reliance on the individual physician's subjective diagnosis based on images of rare cardiac diseases.

Implication of all the available evidence

Integrating effective spatio-temporal DL modeling in real-time cardiovascular imaging studies will increase clinical relevance of AI in assisting non-expert imaging readers for urgently needed triage and management decisions in acute cardiovascular disorders.

1. Introduction

Despite distinct pathogenesis, [1,2] Takotsubo syndrome (TTS) can mimic clinical and electrocardiographic (ECG) features of acute myocardial infarction (AMI), including anterior wall ST segment elevation myocardial infarction (STEMI). Current guidelines advocate the use of coronary angiography to direct differential diagnosis and treatment. [3] Because a substantial portion of TTS cases are actually triggered by bleeding disorders, particularly from the central nervous system, frontline clinicians often face a dilemma when anticoagulation (for cardiac catheterization) or thrombolysis can cause adverse, potentially lethal consequences. Meanwhile, misdiagnosing TTS as STEMI can lead to harmful pharmacological or device-based treatment and worsen hemodynamic compromise. [4,5] During the COVID-19 pandemic, TTS was increasingly found in patients with ECG features of STEMI. [6] For provider protection and capacity leverage, the updated guideline requires point-of-care ultrasound (POCUS) or bedside echocardiography to triage STEMI patients suspected for COVID infection before cardiac catheterization. [7]

TTS-induced myocardial contractile dysfunction usually extends beyond a single (culprit) coronary artery territory. Nonetheless, tethering of nonischemic myocardium adjacent to ischemic or infarcted myocardium often causes two-dimensional (2D) echocardiographic analyses of regional myocardial contractile dysfunction overestimate the actual ischemic region size. Coronary artery anatomic variations also create factual difficulties to distinguish TTS from anterior wall STEMI based on regional wall motion characteristics in bedside echocardiograms.

In daily practice, if clinical characteristics, biomarkers, and ECGs are inadequate for definitive diagnosis, we often have to rely on echocardiography readers' "judgment calls" to support urgent decision-making. In the present study, we investigated whether deep learning (DL) neural networks could reduce erroneous "judgement calls" in the differential diagnosis of TTS and STEMI based on bedside echocardiographic images and videos, and the role of DL in supporting triage and management of cardiovascular emergencies.

2. Methods

2.1. Overview

We trained three deep convolution neural networks (DCNN) and one recurrent neural network (RNN) based on an echocardiographic database from 540 patients in a 10-year retrospective cohort (1 January 2008 to 1 January 2018) at the University of Iowa (UI) and eight other university-affiliated or regional medical centers (Washington University in St Louis, University of North Carolina, State University of New York, Weill Cornell Medical College, Kansas University, Lankenau Medical Center, Northwest Health Medical Center, and Providence Regional Medical Center) inside the United States. An overview of the study design and datasets are illustrated in Fig. 1. The research protocols and waiver of informed consent were approved by the human subjects committee of the UI institutional review board.

2.2. Clinical diagnosis and imaging studies

We obtained clinical, laboratory (Table 1), ECG, angiographic and echocardiographic imaging data of studied patients and followed updated diagnostic criteria for STEMI [8] and TTS. [3] The differentiation between anterior wall STEMI and TTS were all confirmed by coronary angiography (CAG). Cardiac catheterization with selective CAG, left ventriculography (LVG) and percutaneous coronary intervention were performed using standard techniques according to the updated European Society of Cardiology/American College of Cardiology guidelines. [8] Based on coronary artery anatomies, all ventricular segments were divided into culprit and non-culprit artery territories. Two interventional cardiologists, blind to clinical findings, independently evaluated CAG and LVG images. Transthoracic echocardiography (TTE) was performed using standard techniques of 2D echocardiography following the guidelines of the American Society of Echocardiography. [9] All images were stored digitally for playback and subsequent offline analysis. The 2D grayscale images were acquired in the standard apical views, and the standard apical 4-chamber left ventricular (LV) focused view images and videos were used for subsequent studies. Pixel data from the picture archiving and communication systems were preprocessed into numeric arrays, and the data were stored at a resolution of 800×600 pixels. If necessary, they were rescaled through bilinear interpolation. In STEMI patients with cardiac catheterization/coronary angiographically (CAG)-proven significant stenosis ($>70\%$) of the left anterior descending artery (LAD), transthoracic echocardiograms were performed within 24 h of STEMI. Patients were excluded if they had primary valvular disorders, significant pulmonary hypertension, atrial fibrillation, anomalous LAD origin, or no wall motion abnormality in left ventriculography (LVG) and transthoracic echocardiography. Based on anatomic features, Gensini score, and culprit artery location and dominant/major side-branch circulations, segments were divided into culprit or non-culprit artery-supplied areas based on the standard 17-segment LV model and previous publications. [10,11]
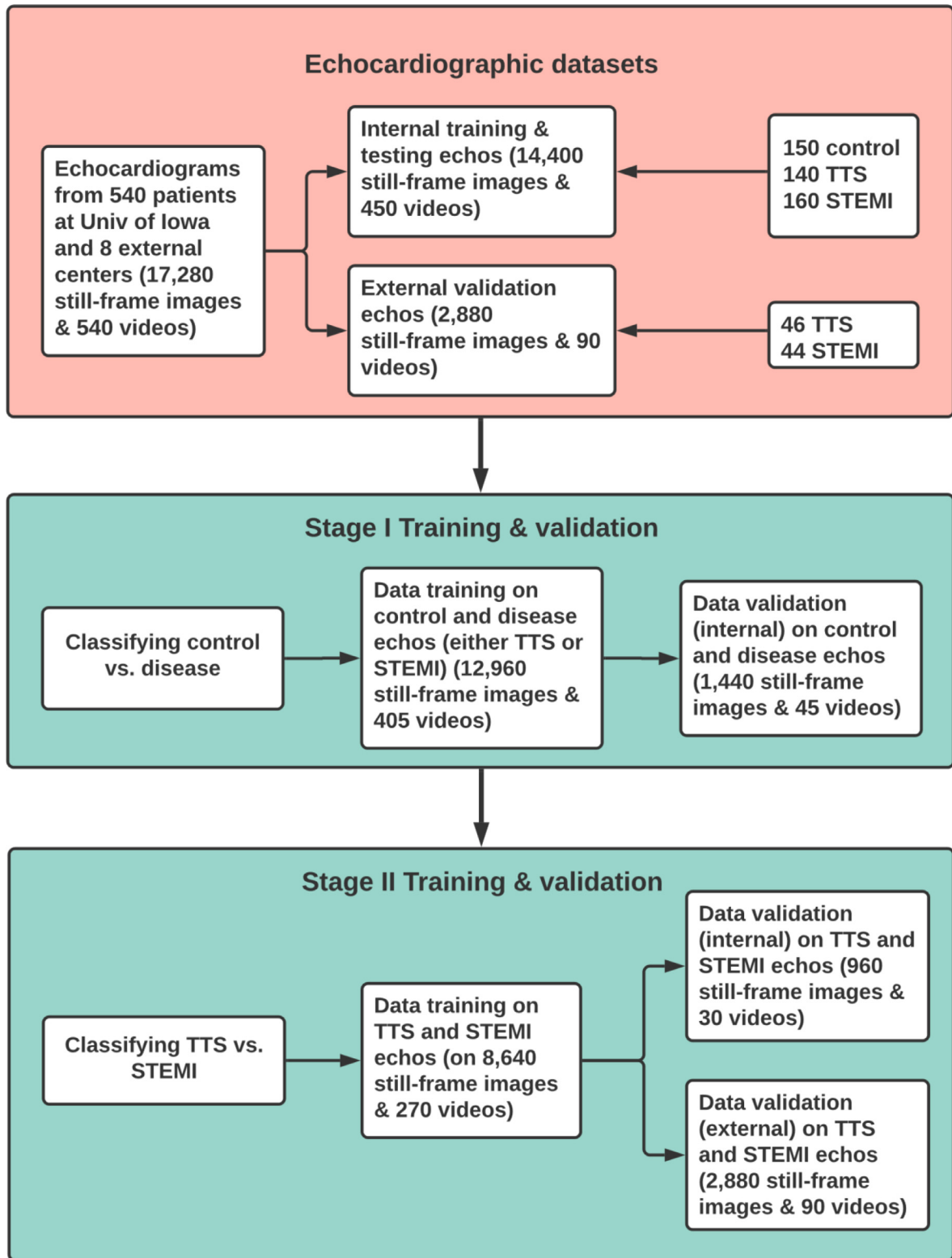
Echocardiographic datasets

Echocardiograms from 540 patients at Univ of Iowa and 8 external centers (17,280 still-frame images & 540 videos)

Internal training & testing echos (14,400 still-frame images & 450 videos)

150 control 140 TTS 160 STEMI

External validation echos (2,880 still-frame images & 90 videos)

46 TTS 44 STEMI

Stage I Training & validation

Classifying control vs. disease

Data training on control and disease echos (either TTS or STEMI) (12,960 still-frame images & 405 videos)

Data validation (internal) on control and disease echos (1,440 still-frame images & 45 videos)

Stage II Training & validation

Classifying TTS vs. STEMI

Data training on TTS and STEMI echos (on 8,640 still-frame images & 270 videos)

Data validation (internal) on TTS and STEMI echos (960 still-frame images & 30 videos)

Data validation (external) on TTS and STEMI echos (2,880 still-frame images & 90 videos)

Fig. 1. Study data for Deep learning (DL) neural network model development.
Ten-Fold cross validation is performed for DL model training and validation by randomly dividing dataset to ten equal subset in each stage. The total number of still frames and videos of each of the folds in different stages of training and validation are shown.

2.3. DL neural networks

2.3.1. Model construction

We developed an ROI selection algorithm (Supplemental methods and Fig. 2A and 2B) to define the regions of interest (ROIs) in the echocardiograms as the input to the DL models, which remove artifacts and labels in the videos and also reduce the computational demands. Three DCNN models and one RNN model were implemented. The two DCNN models (Fig. 2.C.a) were both based on a VGG network, [12] which consisted of nine convolution layers with 3×3 kernel size for each layer of convolution. The convolution stride was fixed to one voxel and the spatial padding for each convolution layer
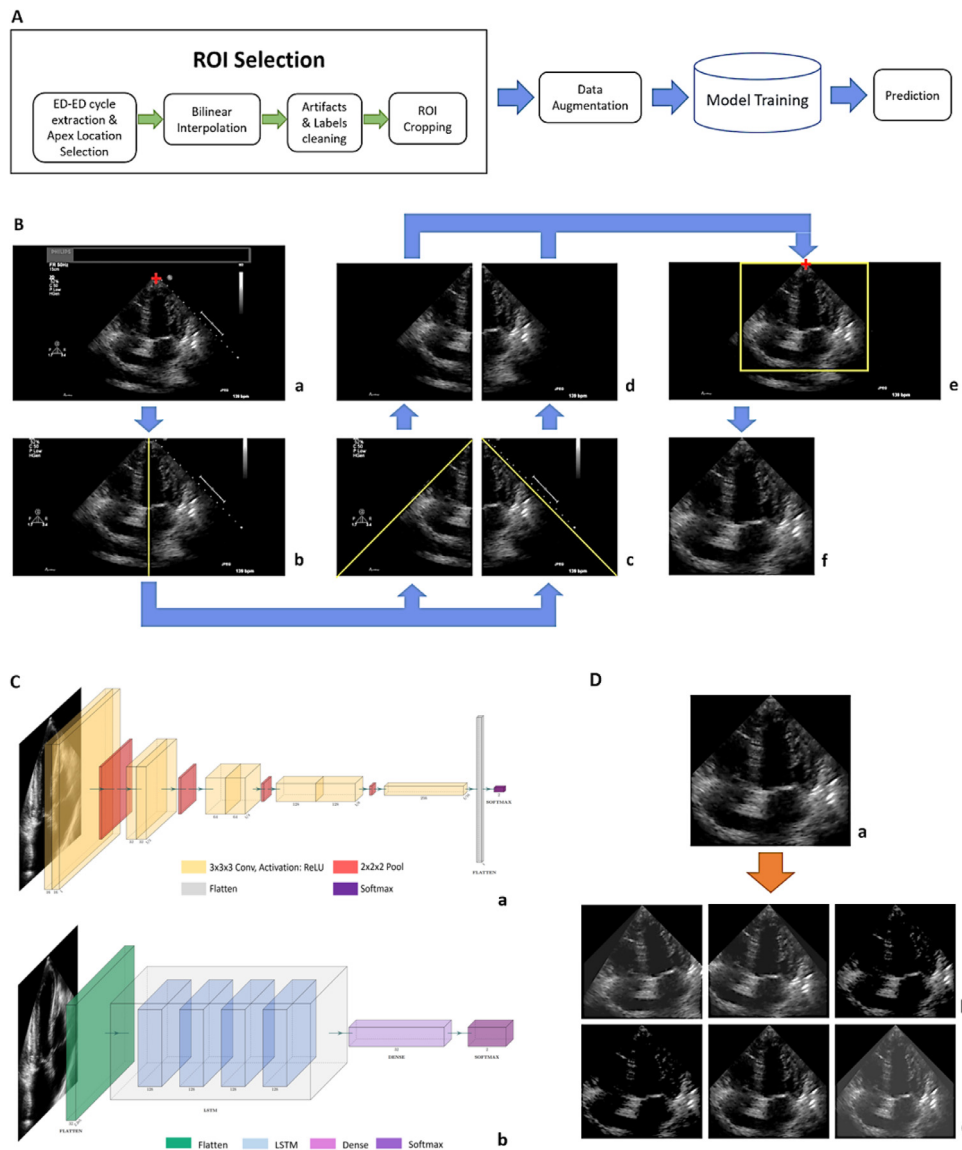
Fig. 2. DL Neural networks.

A. Method workflow for ROI selection, data augmentation, model training and prediction.

B. ROI extraction from the original echocardiogram: Center frame of a cardiac cycle is shown for visual convenience. a. Manual selection of lowest apex location. b. Cropping outside the selected point. c. Left and right part extraction. d. Triangle cut to remove artifacts and labels. e. Joining left and right cleaned parts horizontally. f. Final ROI cut with a (448,480) bounding vox having the selected point in (a) as top middle point of each frame.

C. a. DCNN architecture. b. RNN architecture

D. Data augmentation from selected ROI in (4.a). Top row (4.b) shows augmentation with random rotation and intensity change and bottom row (4.c) shows augmentation with random scaling and intensity change.

input was applied so that the spatial resolution was preserved after convolution. Max-pooling was performed over a 2×2 window to down-sample the feature maps by two, with a stride of one. We started with 16 feature maps in the first convolution layer, which were doubled after every two convolution layers. At the end of these convolution layers, all feature maps were flattened and a dense layer with two channels (one for each class) was added with a soft-max activation to have a probability prediction for each class. All hidden layers were equipped with non-linear rectification (ReLU). [13] In our first DCNN model, we labeled each grayscale frame from all echocardiograms as separate individual cases and used them as the input to the DCNN model with a single-channel (DCNN [2D SCI]). We implemented the second model with exactly the same architecture as the first, but took all frames from a single echocardiogram as a whole to feed into the

DCNN model with multiple-channels (DCNN[2D MCI]). The third model was also based on a VGG network with the same structure as the first two. However, instead of using 2D frames as input, we used the echocardiogram videos as a 3-dimension input. Hence, in this model, all convolution and pooling layers were equipped with $3 \times 3 \times 3$ and $2 \times 2 \times 2$ kernels, respectively. The other network parameters were the same as those in the first two models. We denoted this model as DCNN(2D+t). The fourth model was a recurrent neural network (RNN) with four long short-term memory (LSTM) layers stacked consecutively. The last LSTM layer connected to a dense layer with 32 neurons, followed by a soft-max layer with two neurons for class prediction. We flattened all echocardiogram frames of a video and used them as an input to the recurrent neural network. The DCNN and RNN architectures/algorithms detailed in Fig. 2C and Supplemental methods.

Table 1

Demographic, clinical, and basic echocardiography assessment data of STEMI and TTS patients.

	TTS	STEMI	P-Value
Total Patients, n	140	160	
Age, years (SD)	64·0 (13·1)	61·3 (13·3)	0·08
Male, n (%)	24 (17·1)	112 (70·0)	<0·0001
CAD with PCI, n (%)	16 (11·4)	26 (16·3)	0·27
HTN, n (%)	93 (66·4)	104 (65·0)	0·88
HLD, n (%)	46 (32·9)	77 (48·1)	0·04
DM, n (%)	34 (24·3)	38 (23·8)	0·92
COPD, n (%)	24 (17·1)	10 (6·3)	0·01
CKD, n (%)	21 (15·0)	19 (11·9)	0·46
Hyperthyroid, n (%)	3 (2·1)	1 (0·63)	0·26
Smoking, n (%)[a]	88 (65·7)	107 (67·7)	0·57
ACEi/ARB, n (%)	44 (31·4)	53 (33·1)	0·80
BB, n (%)	51 (36·4)	46 (28·8)	0·24
Chest Pain, n (%)	36 (25·7)	122 (76·3)	<0·0001
Peak CTnT, ng/ml (SD)[b]	0·61 (0·83)	5·63 (7·94)	<0·0001
LVH, n (%)	36 (25·7)	44 (27·5)	0·77
Ejection Fraction <40, n (%)	113 (80·7)	106 (66·3)	0·14
LA Enlargement, n (%)	45 (32·1)	51 (31·9)	0·97

[a] TTS: 134/140 patients; STEMI: 158/160 patients.

[b] TTS: 137/140 patients; STEMI: 147/160 patients.

CAD: coronary artery disease; PCI: percutaneous coronary intervention; HTN: hypertension; HLD: hyperlipidemia; DM: diabetes mellitus; COPD: chronic obstructive pulmonary disease; CKD: chronic kidney disease; ACEi: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; BB: beta blocker; CTnT: cardiac troponin; LVH: left ventricular hypertrophy; LA: left atrium

There were more male STEMI patients than TTS patients (70% vs.17.1%, $P<0.0001$). There were more STEMI patients with hyperlipidemia (48.1% vs. 32.9%, $P = 0.04$). There were more TTS patients with chronic obstructive pulmonary disease (7.1% vs. 6.3%, $P = 0.01$). The mean peak troponin T levels in STEMI subjects was higher than TTS patients (5.63 ng/ml vs. 0.61 ng/ml, $P<0.0001$). More STEMI patients presented with chest pain than TTS patients ($P<0.001$).

2.3.2. Data training and validation

The data training and validation was based on an image database consisting of 17,280 still-frame images and 540 videos from apical 4-chambal view 2D echocardiograms in 540 patients in the University of Iowa (UI) and eight other medical centers. The internal training and validation were performed in two stepwise stages: control versus disease and TTS versus STEMI. We used ten-fold cross-validation for training and validation on 14,400 still-frame images and 450 videos from the echocardiograms of UI patients (150 control, 140 TTS and 160 STEMI). The dataset was randomly divided into ten subsets with the same ratio among the classes as the original dataset to maintain the class balance. Each model was validated on each subset of the data with the model trained on the remaining nine subsets. The performance of the model was the average of all ten validation scores. In addition, each model was trained with augmented data (detailed in Supplemental methods). The models for the TTS vs STEMI classification task were also tested for generalizability on an external dataset consisting of 2880 still-frame images and 90 videos from the echocardiograms of 90 patients with either TTS or STEMI in eight external centers (Fig. 1).

2.4. Human image survey

We used Qualtrics® software (October 2020 version) to create video image surveys: de-identified echocardiographic videos of standard apical 4-chamber view were used for image surveys from 300 individual patients (160 STEMI and 140 TTS). The survey was anonymous and distributed through electronic links to all 57 participants. The only additional information we requested was participants' clinical specialty with training/working time. A total of 49 readers eventually completed all 300 video readings for the human-side classification. They included 30 board-certified cardiologists (8

interventional board-certified cardiologists and 22 National Board of Echocardiography board-certified general cardiologists), 11 senior The American Registry for Diagnostic Medical Sonography board-certified cardiology sonographers, and 8 frontline care (emergency and critical care) physicians with more than three years' experience of POCUS training (Acknowledgement list). The readers were blind to any additional clinical, laboratory, ECG, echocardiography, angiography, or ventriculography data.

2.5. DL and human result comparison

We evaluated and compared image survey results with the cross-validated results of the DCNN(2D+t) and RNN models. We combined all (49) human outputs by majority voting and defined them as human (voting) results. The correctness for human results was defined as the percentage of human readers who made the same diagnosis as the coronary angiography. Conversely, the correctness of the DCNN and RNN models were the estimated probability that the model made the same prediction as the coronary angiography. We examined and reported the data distributions and confusion matrices of the accuracy of the human results in comparison to that of the DCNN(2D+t) and RNN results, and visualized the results using Principal Component Analysis (PCA) method.

2.6. Statistical analysis

We evaluated the performance of the DL neural networks and the human readers using receiver operating characteristic (ROC) curve analysis and confusion matrix with respect to coronary angiography results. Pairwise comparisons of the area under the ROC curve (AUC) were carried out according to the DeLong method [14] while the pairwise comparison of the confusion matrices were applied based on Fisher's exact test. All statistical analysis was performed using the opensource software Python 3.7.4 with package Scipy. Statistical significance was defined as P value <0.05.

2.7. Access to the study data

Drs. FZ, RP, KL and XDW have full access to the study data.

2.8. Role of the funding source

The funders had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the article for publication.

3. Results

The demographic, clinical, and basic echocardiography assessment data of STEMI and TTS patients are summarized in Table 1.

3.1. Visual interpretation

To perform quality assessment of disease prediction, we use the interpretability method of Gradient-weighted Class Activation Mapping (GradCAM), [15] which aims to unfold the activations of the network layers in a deep neural network. In the heatmap, a brighter point indicates that the corresponding pixel in the input image and plays a more important role in class prediction. Fig. 3 showed weighted GradCAM heatmaps overlay on randomly chosen samples from each of the classes for the TTS versus MI classification and the five best- weighted activation maps for each of the randomly chosen TTS or MI samples. The color range of each of the heatmaps are from dark blue to dark red, where dark blue marks the least important and dark red marks the most important pixels for model prediction.
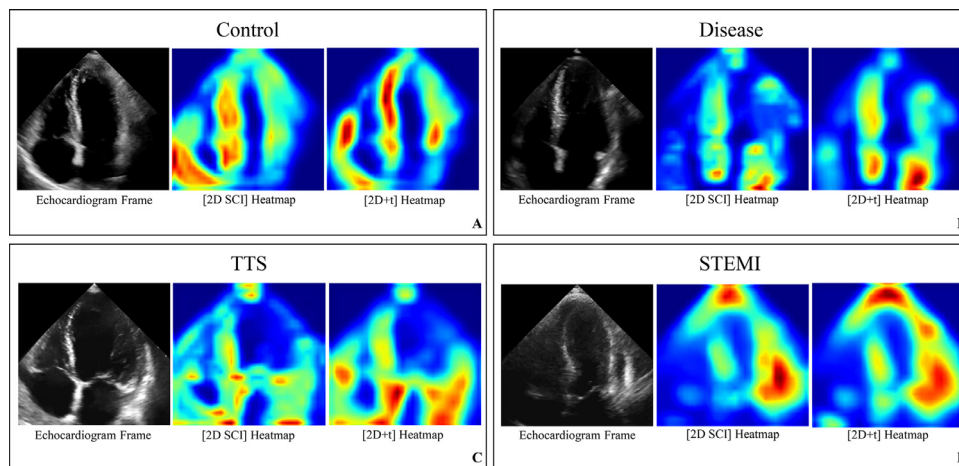
Fig. 3. GradCAM interpretation heatmaps for the prediction with different DCNN models on the Control vs Disease and Takotsubo syndrome (TTS) vs anterior wall ST segment elevation myocardial infarction (STEMI).

A,B,C and D shows the echocardiographic imges with corresponding GradCAM heatmap of DCNN (2D SCI) and DCNN (2D+*t*) prediction models (left to right: Control, Disease, TTS and STEMI, respectively).

Table 2
Prediction accuracy of DL neural networks.

UI Dataset (450 echocardiograms)	Control vs Disease	Model	Control				Disease				Accuracy
			Sensitivity	Specificity	PPV	F1	Sensitivity	Specificity	PPV	F1	
		DCNN (2D [SCI])	0·60	0·87	0·70	0·64	0·87	0·60	0·81	0·84	0·78
		DCNN (2D [MCI])	0·69	0·90	0·79	0·73	0·90	0·69	0·86	0·88	0·83
		DCNN (2D+*t*)	0·86	0·94	0·89	0·87	0·94	0·86	0·93	0·94	0·92
		RNN	0·69	0·87	0·74	0·70	0·87	0·69	0·85	0·86	0·81
UI Dataset (300 echocardiograms)	TTS vs STEMI	Model	TTS				STEMI				Accuracy
			Sensitivity	Specificity	PPV	F1	Sensitivity	Specificity	PPV	F1	
		DCNN (2D [SCI])	0·67	0·78	0·74	0·69	0·78	0·67	0·74	0·75	0·73
		DCNN (2D [MCI])	0·73	0·77	0·74	0·73	0·77	0·73	0·77	0·76	0·75
		DCNN (2D+*t*)	0·79	0·80	0·78	0·78	0·80	0·79	0·83	0·81	0·80
		RNN	0·71	0·79	0·75	0·72	0·79	0·71	0·76	0·77	0·75
External Dataset (90 echocardiograms)	TTS vs STEMI	Model	TTS				STEMI				Accuracy
			Sensitivity	Specificity	PPV	F1	Sensitivity	Specificity	PPV	F1	
		DCNN (2D [SCI])	0·85	0·63	0·72	0·78	0·63	0·85	0·79	0·70	0·74
		DCNN (2D [MCI])	0·94	0·53	0·68	0·79	0·53	0·94	0·92	0·66	0·74
		DCNN (2D+*t*)	0·97	0·57	0·70	0·81	0·57	0·97	0·96	0·70	0·77
		RNN	0·88	0·57	0·69	0·77	0·57	0·88	0·85	0·66	0·73

3.2. Model performance

The class specific sensitivity, specificity, PPV, and F1 scores for different models were shown in Table 2. Briefly, the DCNN(2D SCI), DCNN(2D MCI), DCNN(2D+*t*), and RNN models for the control vs disease prediction showed mean accuracies of 78%, 83%, 92%, and 81%, respectively. The DCNN(2D SCI), DCNN(2D MCI), DCNN(2D+*t*), and RNN models for TTS vs STEMI prediction showed mean accuracies 73%, 75%, 80%, and 75% respectively, and the mean accuracies of 74%, 74%, 77%, and 73% respectively on the external validation.

3.3. Performance comparison/data visualization

The numbers of correct human readings (consistent with coronary angiographic results) on STEMI showed a (left skewed) normal distribution pattern. In contrast, the correct human readings on TTS were rather random (Fig. 4). The confusion matrices showed that the DCNN(2D+*t*) (81·4% correctness) and RNN (70·7% correctness) outperformed human readers (54·3% correctness) to diagnose TTS ($P = 0.000,002$ and 0·006, respectively, Fisher's exact test) while their performances (77·5% and 78·8% correctness) are comparable to that of the human readers (79·4%) on STEMI ($P = 0.786$ and 1·0, respectively, Fisher's exact test) (Fig. 4). The AUC analysis showed that DCNN(2D+*t*) (0·787 vs. 0·699, $P = 0.015$) and RNN models (0·774 vs.

0·699, $P = 0.033$) consistently outperformed human readers in differentiating TTS and STEMI (Fig. 5). In PCA, the DCNN(2D+*t*) result appeared to be the closest to the coronary angiography results, followed by the RNN and the human results (Fig. 5).

4. Discussion

The present study shows that spatio-temporal hybrid DL neural networks can reduce erroneous human "judgement calls" in distinguishing TTS from anterior wall STEMI based on bedside echocardiographic videos, and demonstrates the potential of DL to assist in frontline triage and management of cardiovascular emergencies.

Although echocardiographic videos indeed hold comprehensive imaging information allowing wide-ranging measurements of overt and covert ventricular function, human assessment often subconsciously limits the sampling of spatio-temporal information due to time restriction. While human brains are wired to have a bias toward specific areas of the images based on personal knowledge and experience, DL neural networks speedily analyze every individual pixel, generating the potential to objectively identify delicate features and uncover the predictive ability that may be lost by human readers. [16,17] This lays the foundation for using spatio-temporal convolutions on classification of echocardiographic studies to support real-time differential diagnosis during acute cardiovascular disorders.
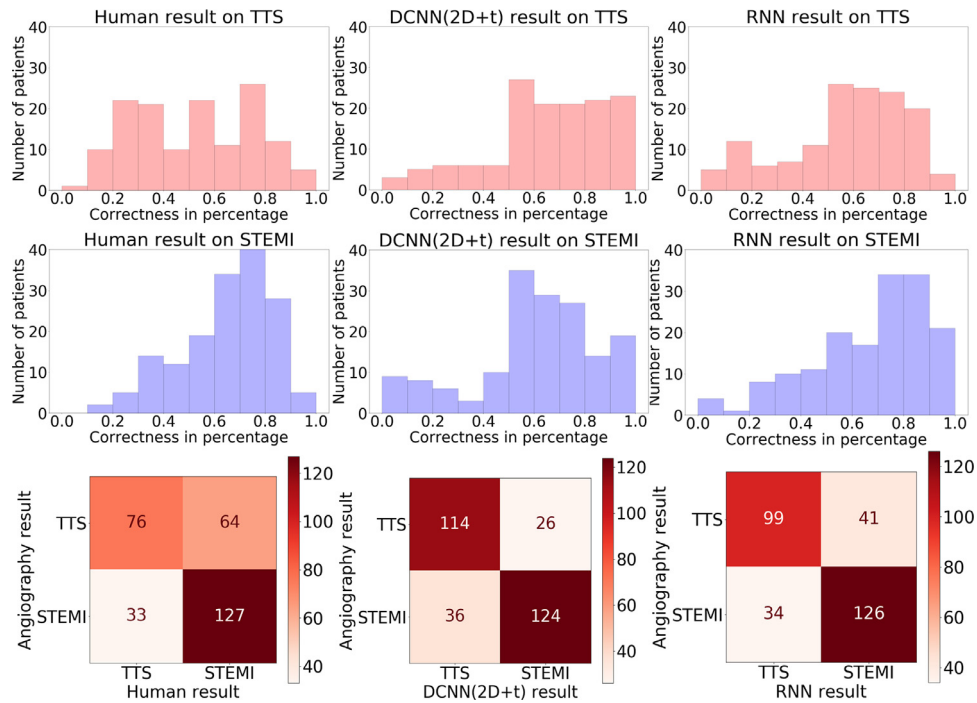
Fig. 4. DCNN (2D+*t*) and RNN reduce erroneous human "judgement calls" on TTS based on bedside echocardiograms.

Top row: histograms of correctness on TTS. Middle row: histograms of correctness on STEMI. Bottom row: confusion matrices.

Left column: human result. Middle column: DCNN(2D+*t*) result. Right column: RNN result.

Correctness for human result is the percentage of human readers who make the same diagnosis as the coronary angiography. Correctness for DCNN and RNN models are the estimated probability the model makes the same prediction as the coronary angiography.
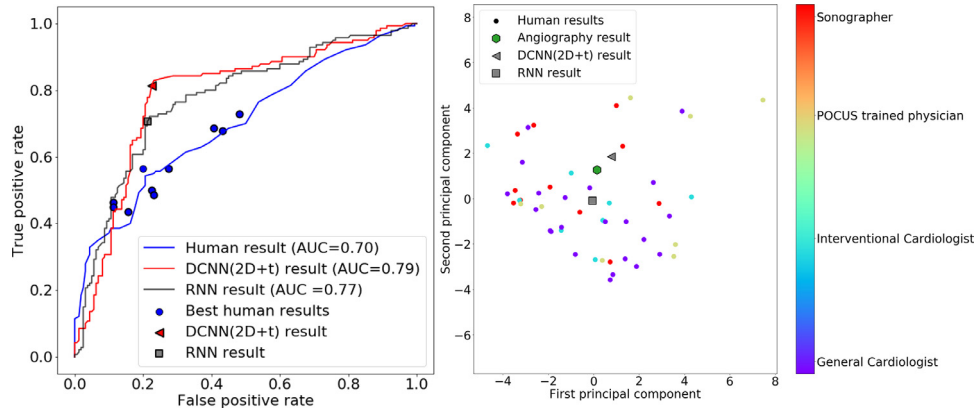
Fig. 5. DCNN(2D+*t*) and RNN models outperform human readers in differentiating TTS and STEMI.

Left: the ROC curves of the overall human, DCNN(2D+*t*) and RNN, as well as the ten best human results. The *P* value of the area under the curve (AUC) between DCNN and the overall human results is 0.015, and the P-value of the AUC between RNN and human result is 0.033. The blue circles represent 10 human readers with best performance.

Right: visualization of DCNN(2D+*t*)/RNN and individual human results with principal component analysis. The human, angiography, DCNN(2D+*t*) and RNN results are shown by different shapes. The specalities of human reader are shown by different colors. Each point represents the comprehensive diagnosis on all 300 echocardiograms from patients (with TTS or STEMI) given by one human reader or one model. These results are projected onto a two-dimensional space using a dimension reduction method – principal component analysis.

DL neural networks have been increasingly used to investigate cardiac pathology based on still-frame echocardiographic images. Although increased image data yield helps binary classifications and computable decision boundaries, their accuracies to identify "advanced" markers appear to be variable. [18−20] The myocardial contractile and relaxing process is highly heterogeneous and time-dependent. Classifying individual image frames in isolation may limit the perception of temporal features during or between cardiac cycles. Instead, composed spatio-temporal information within or between consecutive static images likely empowers the applicability of DL neural models in recognizing subtle changes of myocardial contractility and function. [21 23]

In the present study, DCNN(2D+*t*) model based on echocardiography videos outperformed the DCNN(2D) models based on static image frames. From heatmap data visualization, much of the benefit appeared to be through improved discrimination between certain pairs of views that the DCNN(2D) models found challenging. In such cases, the temporal arm's saliency map showed intense focal activation over the basal LV and right ventricle (RV) segments in echocardiographic videos of TTS and STEMI patients (Fig. 3). The DCNN(2D MCI) model showed slightly improved performance over DCNN(2D SCI), even just with its very limited capacity of making use of spatial feature changes between image frames. The true interpretation emerged when a temporal imaging sequence is integrated. Both

DCNN(2D+t) and RNN models explore temporal motion features in an echocardiographic video. By leveraging spatial and temporal information from multiple image frames across an echocardiographic video, DCNN(2D+t) and RNN models have the potential to detect subtle functional/motion changes through a cumulative evaluation of the continuous movements of the heart, therefore being likely more sensitive than DCNN(2D) models. [21,23] The visualization data support the theory that the DCNN(2D+t) network's ability to discriminate between such classes may be in part due to its ability to track the movement of cardiac structures (basal LV and RV walls) throughout the simultaneous multi-dimensional motion, in order to increase data resolution and catch "invisible" spatio-temporal imaging information. With a limited view of the receptive field, DCNN(2D+t) is able to "see" only a limited number of frames concurrently with the 3D convolution operations, which enables learning of relatively regional ventricular motions. [21,22]

In the present study, DL networks significantly reduced erroneous human "judgement calls" on TTS. In contrast to a normal distribution pattern of the numbers of readers making the correct diagnosis on STEMI, the distribution of the numbers of readers that correctly identified TTS appears to be random and arbitrary (Fig. 4), which did not reflect readers' various training background and experiences. The readers' presumption on low prevalence of TTS in their previous experience may generate judgment bias. In real life practice, it may be further augmented by a fear of missing the diagnosis of possibly life-threatening anterior wall STEMI. Imaging training and research on rare diseases usually relies on establishing national/international registries to build a large-scale image database. Due to a paucity of automated resources for processing raw images and no consistent reporting of data quality measures, this practice requires the great collaboration of multiple medical centers in sharing of data and study model settings. [2] For many such diseases, including TTS, there currently exists no publicly available image database to enrich readers' training and education. Without well-accepted consensus, reading physicians sometimes have to rely on instinct and personal experience to make judgment calls for the diagnosis. The inherent subjectivity likely results in inter-observer variations and errors. Since DL image processing and analysis can reduce human readers' performance variability due to various training background and experience, our results and further data visualization may help explore possible caveats during the imaging training pathways of human readers, which would contribute to tailoring and standardizing future imaging training strategies. [24,25]

The present study has several limitations. 1. Both STEMI and TTS are dynamic and time-variant processes, and contractile dysfunction patterns likely vary at different time points, which has made differential diagnosis even more challenging with a single-time echocardiography. Enriching training echocardiographic datasets with TTS/STEMI imaging features in different evolving stages may help further improve the diagnostic accuracy of DL neural networks. [26,27] 2. The RNN model, which implements the concept of memory with introducing feedback links between layers backward, has the capability of learning temporal context across video frames, which captures more global motion features; however, our current experiments based on limited views of echocardiograms shows that the RNN model is inferior to DCNN (2D+t), which may be due to RNN's relatively weak capability of learning spatial features. An interesting future work would be to unify the strength of both DCNNs and RNNs in a single neural network to more deeply explore the spatio-temporal information to further improve the diagnostic accuracy. 3. The present study was designed to differentiate typical (apical) type TTS and anterior STEMI, which did not apply for atypical (inverse and biventrcilar) type TTS. The diagnostic value of CAG may be limited since TTS can be present even in patients with significant coronary artery disease. [3] Due to the reotrspective nature of the present study, we were unable to apply cardiovascular magnetic resonance

imaging in most patients to define STEMI vs. TTS, which helps further differentiate atypical STEMI and TTS phenotypes. For example, the comparable contractility patterns with typical apical TTS phenotype ("Takotsubo effect") have been increasingly reported and recognized in patients with STEMI, [10,11,28,29] but not been excluded from our training database including echocardiograms in the past 10 years. New training datasets with more delicate phenotyping of STEMI may help further refine the prediction models. 4. Human echocardiography interpretations usually rely on comprehensive echocardiographic techniques in addition to 2D studies, such as Doppler and myocardial strain. The clinical context and information also contribute to the final (differential) diagnosis. Therefore, the present study was not designed to compare the real-life (differential) diagnostic accuracy between AI and human readers. Instead, we aimed to determine the possible added values of DL neural networks to assist non-expert imaging readers for urgently needed disease triage and management decisions during cardiovascular emergencies. During the COVID-19 pandemic, the utilization of comprehensive TTE has been significantly replaced by focused bedside echocardiography and POCUS, to limit exposure and viral transmission. [7] Meanwhile, maintaining diagnostic accuracy and goal-directed therapy in patients with cardiac injuries based on focused echocardiography and POCUS becomes a compelling challenge. Our study serves as a proof-of-concept that DL can streamline and empower currently available bedside imagining tools to effectively and efficiently support real-time triage and management of cardiovascular emergencies, particularly in rural areas or during a global healthcare crisis.

Data sharing statement

Any additional information on methods, research results, extended data, and statements of data availability are available in our submitted supplemental methods and results, and are also available online (UI sharepoint/OneDrive shared folder) with all original echocardiographic images and videos.

Funding

Declaration of Competing Interest

K Liu reports an Obermann Center for Advanced Studies Interdisciplinary Research Grant, University of Iowa, and an ICTS Pilot Grant, University of Iowa.

XD Wu reports an Obermann Center for Advanced Studies Interdisciplinary Research Grant, University of Iowa, and an ICTS Pilot Grant, University of Iowa.

YG Wang declares a National Institutes of Health Award (1R01EB025018−01) and a California State University Dominguez Hills RSCA award to support 3-unit course release in Fall of 2021.

All other authors have nothing to declare.

Acknowledgements

Vandenberg, Dr. Nicholas Walker, Dr. Cory Wittrock, Dr. Josiah Zubairu, Michelle Bader, Ariana Batz, Alyssa Bosko, Madison Coohey, Kristy Crow, Michelle Gilbert, Jill Jordan, Shawndell Leonard, Julie Nguyen, Lucas Rock, and Coral Sickler; Washington University in St. Louis: Dr. Ronald Krone and Dr. Jiafu Ou; State University of New York Upstate: Dr. Hani Kozman, Dr. Jamal Ahmed, and Rob Voelker; Baystate Medical Center: Dr. Mahmoud Abdelghany; University of California − Davis: Dr. Dali Fan; University of Kansas: Dr. Siva Taduru; Cook County Health: Dr. Talal Asif.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.101115.

References

[1] Lyon AR, Citro R, Schneider B, Morel O, Ghadri JR, Templin C, et al. Pathophysiology of Takotsubo Syndrome: JACC State-of-the-Art Review. J Am Coll Cardiol 2021;77(7):902–21.
[2] Templin C, Ghadri JR, Diekmann J, Napp LC, Bataiosu DR, Jaguszewski M, et al. Clinical Features and Outcomes of Takotsubo (Stress) Cardiomyopathy. New England Journal of Medicine 2015;373(10):929–38.
[3] Ghadri J-R, Wittstein IS, Prasad A, Sharkey S, Dote K, Akashi YJ, et al. International Expert Consensus Document on Takotsubo Syndrome (Part I): Clinical Characteristics, Diagnostic Criteria, and Pathophysiology. European Heart Journal 2018;39(22):2032–46.
[4] Liu K, Dogra M. Comatose 62-Year-Old Woman Following Cardiopulmonary Resuscitation. JAMA Cardiology 2018;3(12):1244.
[5] Citro R, Rigo F, D'Andrea A, Ciampi Q, Parodi G, Provenza G, et al. Echocardiographic correlates of acute heart failure, cardiogenic shock, and in-hospital mortality in tako-tsubo cardiomyopathy. JACC Cardiovasc Imaging 2014;7(2):119–29.
[6] Jabri A, Kalra A, Kumar A, Alameh A, Adroja S, Bashir H, et al. Incidence of Stress Cardiomyopathy During the Coronavirus Disease 2019 Pandemic. JAMA Network Open 2020;3(7):e2014780 -e.
[7] Mahmud E, Dauerman HL, Welt FGP, Messenger JC, Rao SV, Grines C, et al. Management of acute myocardial infarction during the COVID-19 pandemic: A Consensus Statement from the Society for Cardiovascular Angiography and Interventions (SCAI), the American College of Cardiology (ACC), and the American College of Emergency Physicians (ACEP). Catheter Cardiovasc Interv 2020;96(2):336–45.
[8] Levine GN, Bates ER, Blankenship JC, Bailey SR, Bittl JA, Cercek B, et al. 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial Infarction: An update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. Catheter Cardiovasc Interv 2016;87(6):1001–19.
[9] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28(1):1–39 e14.
[10] Lei J, Chen J, Dogra M, Gebska MA, Shetty S, Ponnapureddy R, et al. Takotsubo effect" in patients with ST segment elevation myocardial infarction. European Heart Journal: Acute Cardiovascular Care 2020;9(7):711–20.
[11] Qiu Q, Abdelghany M, Subedi R, Scalzetti E, Feiglin D, Wang J, et al. Discrepant myocardial microvascular perfusion and mechanics after acute myocardial infarction: Characterization of the "Tako-tsubo effect" with real-time myocardial perfusion contrast echocardiograph. Int J Cardiol 2019;276:1–7.
[12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv pre-print server. 2015.
[13] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 2012;25:1097–105.
[14] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988:837–45.
[15] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Grad-CAM Batra D. Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision 2020;128(2):336–59.
[16] Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2018;2(3):158–64.
[17] Shah A, Abámoff MD, Wu X. Optimal surface segmentation with convex priors in irregularly sampled space. Med Image Anal 2019;54:63–75.
[18] Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. NPJ Digit Med 2018;1.
[19] Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. npj Digital Medicine 2020;3(1):10.
[20] 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(16):1623–35.
[21] 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.
[22] Huang M-S, Wang C-S, Chiang J-H, Liu P-Y, Tsai W-C. Automated Recognition of Regional Wall Motion Abnormalities Through Deep Neural Network Interpretation of Transthoracic Echocardiography. Circulation 2020;142(16):1510–20.
[23] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature 2020;580(7802):252–6.
[24] Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018;319(13):1317–8.
[25] Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016;375(13):1216–9.
[26] Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist: Reviewed by the American College of Cardiology Healthcare Innovation Council. JACC Cardiovasc Imaging 2020;13(9):2017–35.
[27] Kagiyama N, Piccirilli M, Yanamala N, Shrestha S, Farjo PD, Casaclang-Verzosa G, et al. Machine Learning Assessment of Left Ventricular Diastolic Function Based on Electrocardiographic Features. J Am Coll Cardiol 2020;76(8):930–41.
[28] Sharkey SW, Kalra A, Henry TD, Smith TD, Pink VR, Lesser JR, et al. Coexistence of acute takotsubo syndrome and acute coronary syndrome. Catheter Cardiovasc Interv 2020;96(4):825–9.
[29] D'Onofrio N, Sardu C, Paolisso P, Minicucci F, Gragnano F, Ferraraccio F, et al. MicroRNA-33 and SIRT1 influence the coronary thrombus burden in hyperglycemic STEMI patients. J Cell Physiol 2020;235(2):1438–52.