JACC: CARDIOVASCULAR IMAGING VOL. 14, NO. 2, 2021

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

Artificial Intelligence Aids Cardiac Image Quality Assessment for Improving Precision in Strain Measurements

Kuan-Chih Huang, MD,[a,b] Chiun-Sheng Huang, MD, PhD,[c] Mao-Yuan Su, PhD,[d] Chung-Lieh Hung, MD, PhD,[e] Yi-Chin Ethan Tu, MSIT,[f] Lung-Chun Lin, MD, PhD,[g] Juey-Jen Hwang, MD, PhD[g,h]

ABSTRACT

OBJECTIVES The aim of this study was to develop an artificial intelligence tool to assess echocardiographic image quality objectively.

BACKGROUND Left ventricular global longitudinal strain (LVGLS) has recently been used to monitor cancer therapeutics–related cardiac dysfunction (CTRCD) but image quality limits its reliability.

METHODS A DenseNet-121 convolutional neural network was developed for view identification from an athlete's echocardiographic dataset. To prove the concept that classification confidence (CC) can serve as a quality marker, values of longitudinal strain derived from feature tracking of cardiac magnetic resonance (CMR) imaging and strain analysis of echocardiography were compared. The CC was then applied to patients with breast cancer free from CTRCD to investigate the effects of image quality on the reliability of strain analysis.

RESULTS CC of the apical 4-chamber view (A4C) was significantly correlated with the endocardial border delineation index. CC of A4C >900 significantly predicted a <15% relative difference in longitudinal strain between CMR feature tracking and automated echocardiographic analysis. Echocardiographic studies (n =752) of 102 patients with breast cancer without CTRCD were investigated. The strain analysis showed higher parallel forms, inter-rater, and test-retest reliabilities in patients with CC of A4C >900. During sequential comparisons of automated LVGLS in individual patients, those with CC of A4C >900 had a lower false positive detection rate of CTRCD.

CONCLUSIONS CC of A4C was associated with the reliability of automated LVGLS and could also potentially be used as a filter to select comparable images from sequential echocardiographic studies in individual patients and reduce the false positive detection rate of CTRCD. (J Am Coll Cardiol Img 2021;14:335–45) © 2021 by the American College of Cardiology Foundation.

From the [a]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan; [b]Heart Center, Cheng-Hsin General Hospital, Taipei, Taiwan; [c]Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan; [d]Department of Medical Imaging, National Taiwan University Hospital, Taipei, Taiwan; [e]Division of Cardiology, Department of Internal Medicine, Mackay Memorial Hospital, Taipei, Taiwan; [f]Taiwan AI Labs, Taipei, Taiwan; [g]Section of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; and the [h]Section of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yun-Lin, Taiwan.
The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received December 20, 2019; revised manuscript received July 17, 2020, accepted August 12, 2020.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2020.08.034

ABBREVIATIONS AND ACRONYMS

A2C = apical 2-chamber view

A3C = apical 3-chamber view

A4C = apical 4-chamber view

A4CLS = left ventricular longitudinal strain of apical 4-chamber view

AI = artificial intelligence

CC = classification confidence

CMR = cardiac magnetic resonance

CTRCD = cancer therapeutics–related cardiac dysfunction

ECL = echocardiography core laboratory

LVEF = left ventricular ejection fraction

LVGLS = left ventricular global longitudinal strain

Cancer therapeutics–related cardiac dysfunction (CTRCD) is an increasing problem, and contemporary cancer treatment must not trade cardiac function for cancer remission (1). However, an inappropriate interruption of effective but potentially cardiotoxic cancer treatment may lead to progressive disease or even mortality (2). Therefore, performance expectations with regard to left ventricular ejection fraction (LVEF) are demanding. Left ventricular global longitudinal strain (LVGLS) has been reported to accurately predict a subsequent decrease in LVEF (3,4) and has recently been used in sequential follow-up echocardiography to guide the initiation of cardioprotective therapy (5).

However, image acquisition and analysis are major limitations in the reliability of LVGLS measurements (6,7). Despite improvements in inter-vendor agreement in LVGLS after the European Association of Cardiovascular Imaging/American Society of Echocardiography task force strain standardization initiatives (8) and the development of automated software to further improve inter-rater reliability, speckle tracking analysis still requires a learning curve (9), and the image quality still significantly affects test-retest reliability (10). Furthermore, it is difficult to objectively and quantitatively gauge the concept of image quality because it is subjectively assessed in echocardiography core laboratories (ECLs) (11,12). This not only hinders the democratization of LVGLS but also limits its application in sequential follow-up clinical scenarios. To overcome these obstacles, artificial intelligence (AI) has recently been used to assess echocardiography image quality (13,14).

The aim of this study was first to develop an AI-based assessment of echocardiographic image quality by quantifying video features. We then investigated the predictive value of the AI quality marker on the reliability of automated LVGLS and CTRCD false positive rate.

METHODS

There were 3 populations in this study (**Central Illustration**). In population 1, we trained our AI model to be capable of identifying views by extracted features and derived the parameters of classification as quality markers (Supplemental Appendix). Patients with breast cancer (population 2) were investigated by references of intermodality consistency (population 3) for determining the cutoff of the quality marker, to prove the concept that AI-based quality assessment improves precision in strain measurements.

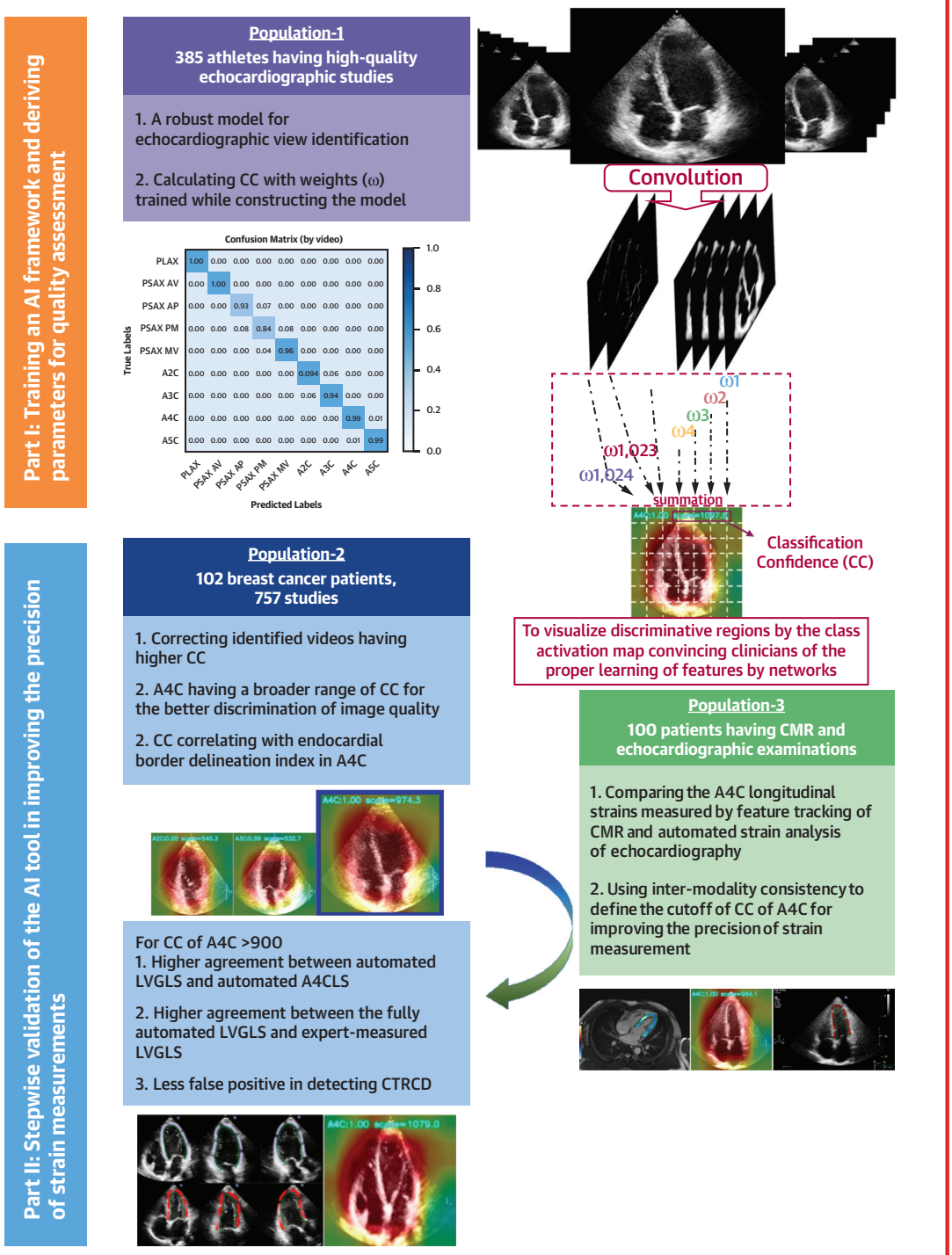
POPULATION 1: TRAINING THE AI MODEL. The Check-up Your Heart Program directed by the International University Sports Federation (Fédération Internationale du Sport Universitaire) at the 2017 Summer Universiade in Taipei was a peri-participation screening program of the cardiovascular system. We enrolled 826 volunteer student-athletes and collected standard echocardiography data recorded by experienced technicians. From the total 826 studies, 385 high-quality studies were randomly assigned as training, validation, and test sets (Supplemental Table S1) to develop a deep learning driven computer vision algorithm for frame-by-frame view identification.

CONVOLUTIONAL NEURAL NETWORKS FOR VIEW IDENTIFICATION. Convolution and classifier are the 2 principal components of neural networks, and they mimic the way a visual system works.

Convolution. Using characteristics of weight sharing and hierarchical connections, the convolution at the beginning layers typically detects simple features such as edges in an input image (**Figure 1A**). It combines simple features to form more complicated abstractions, such as shapes, at the following layers. The final output is an assembly of maps of extracted features for training the classification. Furthermore, the pooling process enhances the robustness of the network with regard to spatial variances. In this study, we used the DenseNet-121 (15) convolutional neural network because it has the advantages of concatenating, reusing, and conserving features with low complexity.

Classifier. The calculation of global average pooling flattened the final stack of convolutions to a vector to train the classification weights through a fully connected layer (**Figure 1B**). In our framework, the outputs of view identification included 9 standard views (parasternal long-axis view, parasternal short-axis view at the aortic valve level, parasternal short-axis view at the mitral valve level, parasternal short-axis views at the papillary muscle level, parasternal short-axis view at the LV apex level, apical 4-chamber view [A4C], apical 2-chamber view [A2C], apical 3-chamber view [A3C], and apical 5-chamber view). Accordingly, the weighted sum of the final vector consisted of 9 numbers of classification confidence (CC) for the corresponding views, which were used by the probability function (softmax) (Supplemental Appendix) to determine the category

CENTRAL ILLUSTRATION Schematic Chart of the Study Process

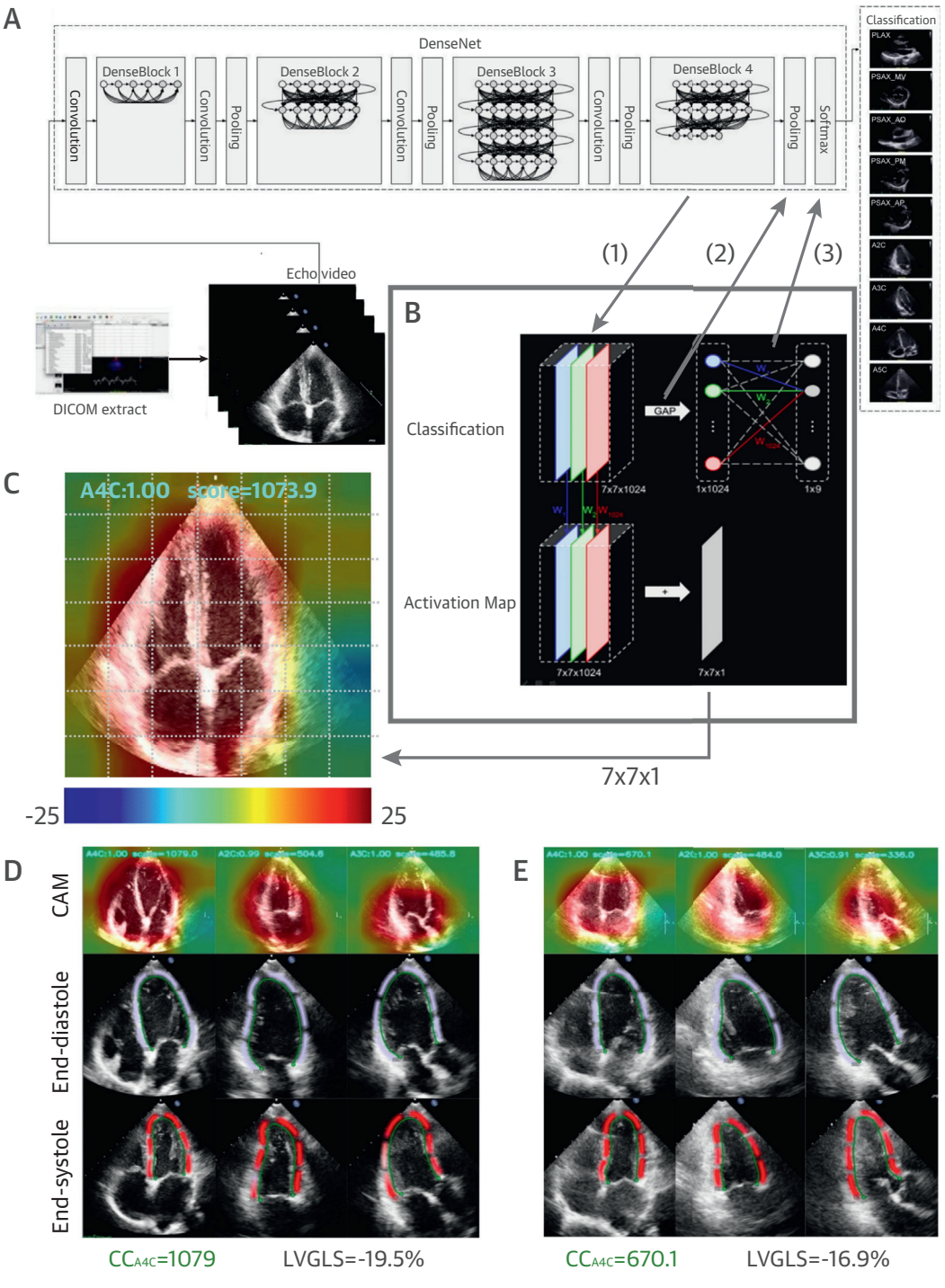


Huang, K.-C. et al. J Am Coll Cardiol Img. 2021;14(2):335–45.

Characteristics of the 3 populations with corresponding main tasks and study results. AI = artificial intelligence; A2C = apical 2-chamber view; A3C = apical 3-chamber view; A4C = apical 4-chamber view; A4CLS = left ventricular longitudinal strain of apical 4-chamber view; CC = classification confidence; CMR = cardiac magnetic resonance; CTRCD = cancer therapeutics–related cardiac dysfunction; LVGLS = left ventricular global longitudinal strain.

JACC: CARDIOVASCULAR IMAGING, VOL. 14, NO. 2, 2021
FEBRUARY 2021:335–45

FIGURE 1 Densely Connected Convolutional Neural Networks for Training the Classification of Echocardiographic Views and Developing the Class Activation Map



Continued on the next page

of maximum probability (**Figure 1B**, (3)). The robust performance of our view identification AI model is shown in Supplemental Figures S1 and S2.

CLASSIFICATION CONFIDENCE AND CLASS ACTIVATION MAP. Zhou et al. (16) demonstrated that the convolutional units of various layers of a neural network could behave as object detectors without supervision for the location of the object. Therefore, the weights trained by the fully connected layer could be backward projected to the feature maps to produce a class activation map (**Figure 1C**) to highlight the class-specific discriminative regions.

Rationally, the more extracted features an input image has, the higher quality it has for an assigned view classification. The softmax function maximizes the largest classifier output (number 9 in our study) for prediction and discards the uncertainty. CC of an assigned view can be thought of as a perception of that view inside the "black box" of neural networks. We also used the class activation map as a visualization tool to convince clinicians that the features generated by networks were appropriate. Finally, we applied the averaged CC of a labeled view as the assessment of image quality for that video.

POPULATION 2: PATIENTS WITH BREAST CANCER WITHOUT CTRCD. We retrospectively collected echocardiograms from patients with breast cancer who were enrolled in a randomized control trial at National Taiwan University Hospital. These patients had >5 sequential follow-up echocardiographic studies and were considered to be free from CTRCD at the end of the trial. A total of 102 patients (age 51.6 ± 9.1 years; body mass index 23.6 ± 3.7 kg/m^2; 48 with left-sided and 54 with right-sided breast cancers; 41 with breast conservation therapy, and 61 who had subtotal mastectomies with 6 reconstructions; none had coronary artery disease) met our inclusion criteria (35 from the APHINITY [A Study of Pertuzumab in Addition to Chemotherapy and Trastuzumab as Adjuvant Therapy in Participants With Human Epidermal Growth Receptor 2 (HER2)-Positive Primary Breast Cancer; NCT01358877] trial (17), 4 from the SafeHer [A Safety and Tolerability Study of Assisted and Self-Administered Subcutaneous (SC) Herceptin (Trastuzumab) as Adjuvant Therapy in Early Human Epidermal Growth Factor Receptor 2 (HER2)-Positive Breast Cancer; NCT01566721] trial (18), 18 from the KATHERINE [A Study of Trastuzumab Emtansine Versus Trastuzumab as Adjuvant Therapy in Patients With HER2-Positive Breast Cancer Who Have Residual Tumor in the Breast or Axillary Lymph Nodes Following Preoperative Therapy; NCT01772472] trial, 26 from the KAITLIN (A Study of Trastuzumab Emtansine (Kadcyla) Plus Pertuzumab (Perjeta) Following Anthracyclines in Comparison With Trastuzumab (Herceptin) Plus Pertuzumab and a Taxane Following Anthracyclines as Adjuvant Therapy in Participants With Operable HER2-Positive Primary Breast Cancer; NCT01966471] trial, and 19 from the KRISTINE [A Study Evaluating Trastuzumab Emtansine Plus Pertuzumab Compared With Chemotherapy Plus Trastuzumab and Pertuzumab for Participants With Human Epidermal Growth Factor Receptor 2 (HER2)-Positive Breast Cancer; NCT02131064] trial (19)). There was no significant change in LVEF during the chemotherapy course (LVEF immediately before the trial was 70.4 ± 4.1% and the last LVEF after the trial was 70.2 ± 5.2%; p = 0.70). Serial follow-up echocardiographic studies (n = 752) of all 102 patients were analyzed for automated LVGLS and CC. Automated strain analysis was conducted using dedicated software (the AutoSTRAIN module, TTA 2.30, TomTec Imaging Systems GmbH, Unterschleissheim, Germany).

Population 3: comparing the LV longitudinal strain of A4C (A4CLS) between 2 different imaging modalities. We retrospectively and consecutively collected 100 patients (67 men; age 51.5 ± 16.8 years) who had both

FIGURE 1 Continued

(A) The Dense Blocks in DenseNet-121 maintain constant dimensions of feature maps. The volume after every dense block grows with concatenating filtered features but is down-sampled in transition. The input matrix (224 × 224) is convoluted and pooled to the last stage of the network with 1,024 dimensions of 7 × 7 matrixes (1). **(B)** The global average pooling (GAP) transforms the final tensor into a 1 × 1,024 vector (2) for training weights through a fully connected layer. Nine artificial intelligence-trained weighted summations corresponding to each classification are analyzed using the softmax function (3) to decide the probability of view identification. We localized the image regions with discriminative importance by projecting back the weights supporting the predicted classification to the corresponding 1,024 feature maps in the last convolution to obtain a summation. **(C)** The class activation map (CAM) in the 7 × 7 matrix was overlaid on the input image and visualized by the JET colormap function with 25 as the saturation threshold. As the CAM example shows, the possibility of this still frame to be classified as an apical 4-chamber view is 1.00 (A4C:1.00) with classification confidence (CC) of 1,073.9 (score = 1,073.9). (D and E) The serial echocardiographic studies of 1 patient on 2 occasions, including automated tracking at end-diastole and end-systole, demonstrated that a lower CC of the apical 4-chamber view (CC_{A4C}) leads to a lower estimated left ventricular global longitudinal strain (LVGLS).

cardiac magnetic resonance (CMR) imaging and transthoracic echocardiograms from November 2018 to December 2019 at our hospital. All of the patients had sinus rhythm, normal LVEFs, and no diagnosis of cardiomyopathy. The interval between these 2 examinations was <30 days and clinically event free. Four-chamber views of the CMR datasets were analyzed using feature tracking (2-dimensional cardiac performance analysis MR, TomTec Imaging Systems GmbH) (20). Automated strain analysis of the corresponding echocardiograms was conducted with the same software, as mentioned previously.

Ethical approval/consent. Fédération Internationale du Sport Universitaire authorized the Taiwan Society of Cardiology to collect data and conduct surveys, and this was approved by the institutional review board of National Taiwan University Hospital (201706040RIND for population 1). All athlete participants provided written informed consent. Retrospective echocardiography analyses for patients with breast cancer and comparisons with CMR were approved by the institutional review board of National Taiwan University Hospital (201902042RIND for population 2; 201809013RINB for population 3).

STATISTICAL ANALYSIS. Categorical variables were compared using the chi-square test, and continuous variables were expressed as mean ± SD. Based on their distributions, continuous variables were compared using Student's *t*-test or Mann-Whitney U test. Spearman's rank correlation method was used as a nonparametric measure of associations between LVGLS obtained from multiple views and A4CLS from a single view. Receiver-operating characteristic curve analysis was used to predict a <15% relative difference in A4CLS derived by CMR and echocardiography. Intraclass correlation coefficients were calculated to assess reproducibility. All statistical analyses were performed using STATA version 13 (StataCorp., College Station, Texas). All reported p values were 2-tailed, and p values <0.05 were considered statistically significant.

RESULTS

PERFORMANCE OF VIEW IDENTIFICATION AI (POPULATION 1). The robust performance of our AI model is shown in Supplemental Figures S1 and S2. The source code and model weights have been uploaded in GitHub (https://github.com/ailabstw/EchocardiographyQC; Supplemental Figure S3 for software workflow).

REAL-WORLD ECHOCARDIOGRAPHY FROM PATIENTS WITH BREAST CANCER (POPULATION 2). We applied the AI tool developed from population 1 to population 2 in the 3 apical views (A4C, A2C, and A3C) that are required to calculate LVGLS. The AI correctly identified 702 (92.7%) A4C, 603 (79.7%) A2C, and 622 (82.2%) A3C videos.

CHARACTERISTICS OF CC IN POPULATION-2. Regarding the 3 apical views, CC was 726 ± 220 for A4C, 403 ± 134 for A2C, and 347 ± 154 for A3C. **Figure 2** illustrates the numbers of videos at different CC thresholds.

CC values were higher in the correctly identified videos than in those that were misclassified (A4C: 763 ± 177 vs. 252 ± 155; $p < 0.05$; A2C: 436 ± 110 vs. 253 ± 133; $p < 0.05$; A3C: 394 ± 121 vs. 162 ± 126; $p < 0.05$).

Endocardial border delineation was classified using a 3-point scoring system (21,22) (0 = not visible; 1 = fairly visible; and 2 = clearly visible in the selected cardiac cycle) in the 6 segments of the LV in the A4C view. The endocardial border delineation index of A4C was defined as the sum of scores. We randomly calculated the endocardial border delineation index of 20% A4C videos in population 2, and found that CC was significantly correlated with endocardial border delineation index (CC = 93.8 × endocardial border delineation index + 23.2; $r = 0.82$; $p < 0.001$).

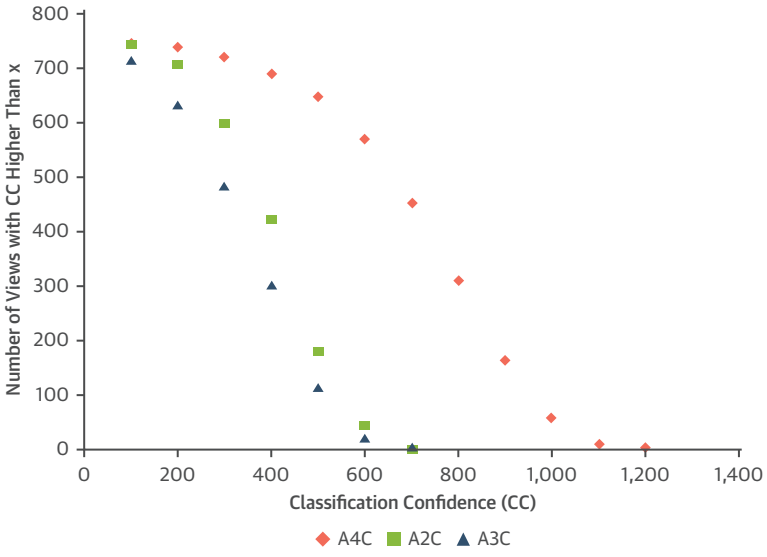
PREDICTION OF IMAGE QUALITY FOR INTERMODALITY CONSISTENCY OF STRAIN MEASUREMENTS (POPULATION 3). We compared the A4CLS derived from CMR feature tracking and automated echocardiographic speckle tracking (Videos 1 to 6). **Figure 3C** illustrates that the relative difference in A4CLS and the CC were significantly correlated with modest linearity ($r = 0.71$; $p < 0.0001$).

Previous studies reported that the limit of agreement for different methods of LVGLS was 3% (23,24); for a normal LVGLS of −20%, this was also a 15% relative difference to define CTRCD (25,26). CC was used in receiver-operating characteristic analysis to predict whether the relative difference in strain values was within 15% between CMR and automated echocardiographic-derived A4CLS (**Figure 3D**), which showed an area under the curve of 0.837 and a CC cutoff point of 900 (sensitivity 48%, specificity 90%).

CUTOFF OF CC FOR A4C VIDEOS AND ITS EFFECTS ON A2C AND A3C IN POPULATION 2. Of the 757 studies in population 2, 164 had the CC >900 for A4C videos. For studies with CC of A4C >900, the A2C (473 ± 122 vs. 384 ± 141; $p < 0.001$) and A3C (400 ± 143 vs. 332 ± 153; $p < 0.001$) videos also had higher CC.

CUTOFF OF CC AND AGREEMENTS OF LVGLS IN POPULATION 2. Parallel forms reliability. A single-view A4CLS might substitute standard LVGLS from 3 apical images because the former efficiently detects

FIGURE 2 **Trends of Numbers of Images With Increasing CC at Every 100 Points for 3 Apical Views**



Note that apical 4-chamber view (A4C) has a broader range of CC than the apical 2-chamber view (A2C) and the apical 3-chamber view (A3C), which indicates A4C has more discriminative regions for the convolutional neural network. Other abbreviation as in Figure 1.

the diffuse process of cardiotoxicity (23). We measured the reliability (parallel forms) between these 2 assessments of LV longitudinal strain. Figure 4 shows a better agreement between automated LVGLS and A4CLS for those patients with CC of A4C >900.

Inter-rater reliability. We randomly selected 20 studies with CC of A4C >900 and 20 studies with CC <900 to compare the agreement between fully automated LVGLS and the expert-measured LVGLS. The intraclass correlation coefficients were 0.689 for CC >900 and 0.067 for CC <900.

CC OF A4C >900 AND THE VALIDITY OF LVGLS. For patients with breast cancer without CTRCD, the reported reference of LVGLS is approximately −20 ± 4% (25). A higher percentage of videos with a CC of A4C >900 had a strain value within this range (LVGLS: CC <900, n = 387 [65%]; CC >900, n = 143 [87%]).

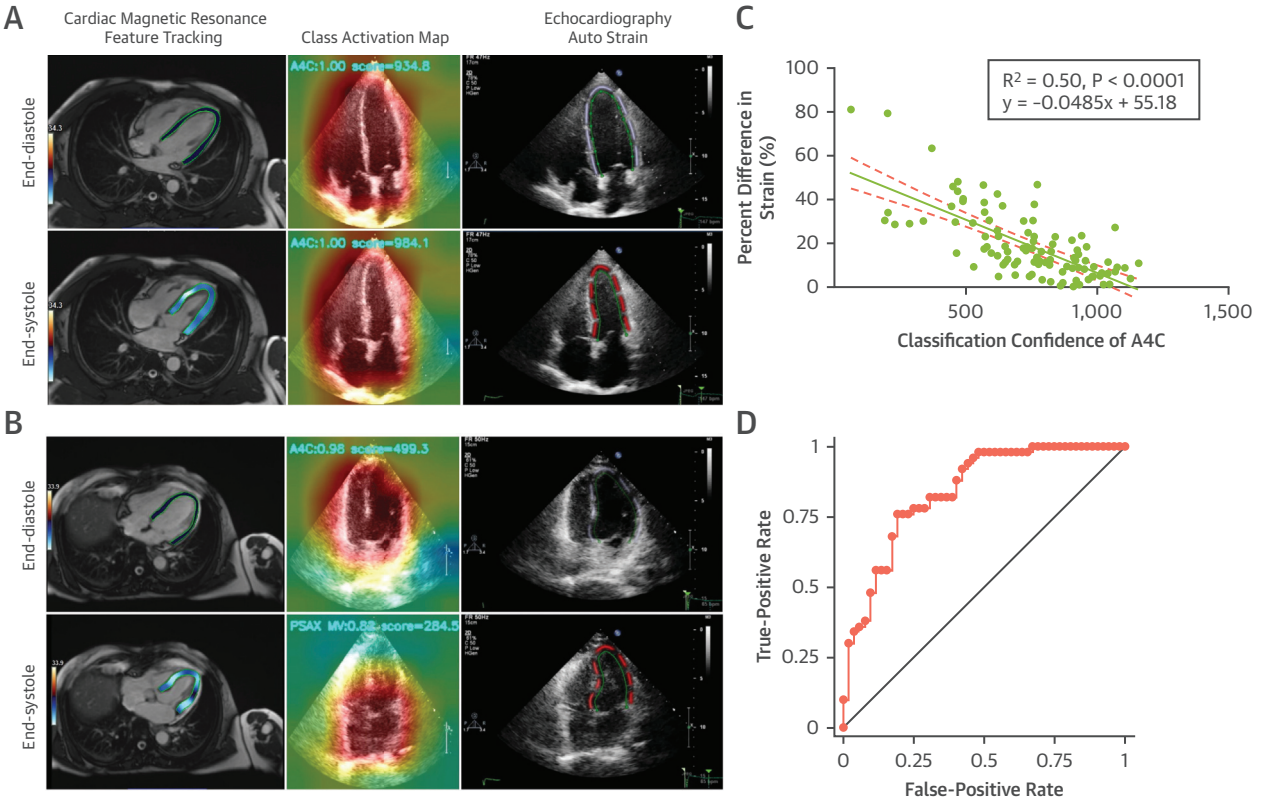
Studies with a CC of A4C >900 had similar LVEFs to those with a CC <900 (71.1 ± 5.0% vs. 70.6 ± 4.9%; CC >900 vs. CC <900; p = 0.227). However, the LVGLS value was nearly 15% lower in those with a CC of A4C <900 (LVGLS: −19.5 ± 2.7% vs. −16.9 ± 3.1%; CC >900 vs. CC <900; p < 0.001).

TEST-RETEST RELIABILITY OF LVGLS AND FALSE POSITIVE CTRCD. Individuals with repeat echocardiographic studies were selected for sequential comparisons of automated LVGLS. In the subgroup with a CC of A4C <900, 96 patients had repeat studies and 44 patients had a 15% relative decrease in LVGLS value. Of those with a CC of A4C >900, 44 patients had repeat studies, but only 9 patients had a 15% decrease in LVGLS value (chi-square = 5.5861; p = 0.018).

DISCUSSION

In this study, we enrolled 3 populations to prove the concept that AI can objectively assess image quality and serve as the first step of an echocardiographic measurement pipeline. We used population 1 to train a view classification AI model. Because CC represents the aggregation of discriminative features learned by the convolution process, our approach to derive an algorithm from a young athletic population was successful because most of these subjects had good images. The algorithm might be particularly applicable to assess the quality of detecting subclinical alterations. Therefore, we collected a cohort of patients with breast cancer (population 2). In that circumstance, the detection of CTRCD by strain measurements was an important issue during serial follow-up. The A4C had a broader range of CC to discern image quality, and the CC significantly correlated with conventional quality parameters. Through the intermodality consistency with the CMR feature tracking (population 3), we determined the CC

FIGURE 3 Comparison of 2 Imaging Modalities to Define the Cutoff of CC of the A4C for Adequate Image Quality



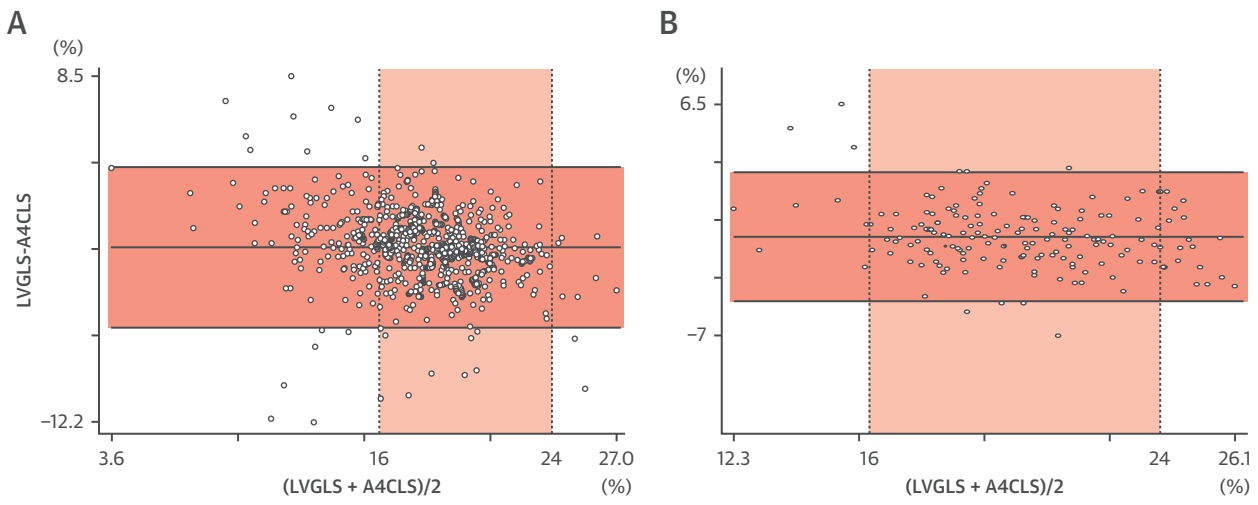
Corresponding cardiac magnetic resonance feature tracking, class activation map, and automated ultrasonic speckle tracking (auto strain) in A4C views are shown at end-diastole and end-systole for the individual patient with echocardiograms in **(A)** sound quality (Videos 1 to 3) and **(B)** poor quality (Videos 4 to 6). Note that the poor quality echocardiography is misclassified into a parasternal short-axis view at the mitral valve level at end-systole **(B, middle, lower panel)**. **(C)** The relative difference (in absolute value) of longitudinal strain measured by echocardiography to cardiac magnetic resonance is plotted versus CC, which demonstrates a significant correlation with modest linearity ($r = 0.71$; $p < 0.0001$). **(D)** Receiver-operating characteristic curve analysis shows that the CC >900 can predict a relative difference in longitudinal strain values of <15% between these 2 modalities (area under the curve [AUC]: 0.837; sensitivity: 48%; specificity: 90%). Abbreviations as in Figures 1 and 2.

cutoff of A4C for improving the precision of echocardiographic strain measurements. Although standard LVGLS remained the preferred technique in CTRCD (23), the better parallel form reliability at CC >900 made the A4CLS a possibly more efficient tool. We finally applied CC on population 2 with repeated echocardiographic examinations. We demonstrated the usefulness of AI tools to prevent false positive detection of CTRCD in serial comparison for an individual patient. Adapting and reverifying the model might be necessary when conducting image quality assessments in other cardiac diseases.

Regarding the use of CC to assess image quality in the detection of CTRCD, the likelihood of a reduced CC due to an actual decline in longitudinal strain raised an important issue. Because no patients had a reduced LVEF in the present study, we collected a population with dilated cardiomyopathy and satisfactory image acquisition (n = 50; LVEF: 32.2 ± 8.2%; LVGLS: −7.0 ± 2.9%; A4CLS: −8.1 ± 3.1%). The CC of A4C videos in this population was 879 ± 142, which was not lower than that of population 2 in our study (726 ± 220). Accordingly, an actual drop in LVGLS did not lead to a decrease in CC in our framework. This might be because more discriminative features could be maintained for a less contractile ventricle.

IMPACT OF CC ON THE SEQUENTIAL FOLLOW-UP OF CARDIAC FUNCTION. Although echocardiography is limited by inter-rater reliability and image quality, its availability means that it is still the most frequently used follow-up tool. Many studies have

FIGURE 4 **Bland-Altman Plots of Agreement Between LVGLS and Left Ventricular A4CLS**



The mean differences ± limits of agreement were −1.75 ± 4.80% for CC of A4C **(A)** <900 and **(B)** −1.23 ± 3.76% for CC >900. The **vertical block** shows the value of normal longitudinal strain (16% to 24% in absolute value). At CC >900, the agreement between LVGLS and the longitudinal strain of apical 4-chamber view **(**A4CLS) is better. As to the value of (LVGLS+A4CLS)/2, videos with a CC >900 are prone to have a longitudinal strain value within the normal range. Abbreviations as in **Figures 1 and 2**.

demonstrated that LVEF is not that objective and that it can sometimes be subjective and profoundly affected by the experience of the sonographers (27). In contrast, even with known variability in measurements, LVGLS has still been shown to have prognostic implications in patients with CTRCD (28). ECLs standardize and ensure the quality of measurements to overcome the operational limitations of echocardiography in clinical investigations. Nevertheless, such dependence on ECLs hinders the democratization of LVGLS. A recent study suggested that a >15% relative reduction in LVGLS from baseline could be considered a marker of early LV subclinical dysfunction (25,26). However, as demonstrated in our results, inadequate image quality could also lead to a >15% deviation in the value of LVGLS. As a result, it was difficult to tell whether this deviation was an actual drop in LVGLS or the effect of compromised image quality. To investigate the effect of image quality on the test-retest reliability of LVGLS, we retrospectively enrolled patients with cancer from 5 randomized controlled trials in whom CTRCD was excluded. Accordingly, we assumed no or negligible temporal variability in the real value of LVGLS as previously reported (29). Therefore, the test-retest reliability of automated A4CLS in our study could be mostly attributed to the image quality. Automated strain analysis software can help to overcome inter-rater reliability, and our results indicated that the test-retest reliability of LVGLS could be improved through an AI-based quality marker tool to screen out comparable images from serial follow-up echocardiographic studies. CC, in combination with automated measurement technology, can thus facilitate the workflow and ameliorate the loads in ECLs, especially for cardiac diseases that need frequent follow-up echocardiograms.

CC OF A4C VIEW COULD SERVE AS A QUALITY INDEX FOR AUTOMATED LVGLS. In the present study, we used the CC of A4C videos as the key factor for assessing the reliability of LVGLS values. This might raise concerns because LVGLS is actually derived from A4C, A2C, and A3C videos. There were 2 main reasons why we chose to use CC of A4C instead of combining the CC of all 3 views via weighted summation or other methods. First, the CC of A4C had a broader range than that of A2C and A3C (**Figure 2**), which implied the CC of A4C had a better discriminative ability as a quality marker. Second, the CC values for A4C, A2C, and A3C were modestly but significantly correlated (Pearson's r value for CC between A4C and A2C was 0.38; $p < 0.001$; for A4C

and A3C, r = 0.19; p < 0.001; and for A2C and A3C, r = 0.33; p < 0.001). Technically, A4C is usually the first acquired image of these 3 apical images, and a foreshortened A4C is often followed by a foreshortened A2C and an A3C that is exactly a low parasternal view. Consequently, examinations with a higher CC of A4C also had significantly higher CC of A2C and A3C.

IMPORTANCE OF IMAGE QUALITY ASSESSMENT IN THE DEVELOPMENT OF ECHOCARDIOGRAPHIC AI. In contrast to other imaging modalities, most of the recently developed AI applications for echocardiography are limited to images of good quality (30-32). The concept of an AI-based quality marker facilitates human–computer interactions and may lead to a transition away from human-based quality control. Our results suggested that the CC of a specific view from a successful classification model could serve as the quality marker for that view. In addition, we used the intermodality concordance in strain measurements to validate the role of CC in evaluating image quality, which proposed an AI pipeline useful to appropriately integrate information from multiple modalities. The assurance of image quality could then accelerate both the development and the external validation of AI analytic tools for echocardiography.

STUDY LIMITATIONS. First, we hypothesized that the clinical conditions were stable when retrospectively comparing LVGLS from sequential follow-up echocardiograms at different time points. We attributed fluctuations or deviations in LVGLS only to the image quality. However, because there were no patients with CTRCD in our study population, further prospective studies are needed to investigate the effects of CC on CTRCD false negative detection rates and strategies for the following management.

Second, only 22% (164 of 757) of the population 2 met the criteria of CC of A4C >900, and we did not collect high-quality images from ECLs in other areas, as in the European Association of Cardiovascular Imaging/American Society of Echocardiography task force strain standardization initiatives. As a result, further prospective studies are needed to investigate the feasibility of setting the criteria for CC as high as >900. Although we can debate that LVGLS was not included in the echocardiographic protocols of the included breast cancer studies at that time, further communication and cooperation between ECLs are still necessary to build a more comprehensive echocardiographic databank to validate an adaptive threshold of the CC for various tasks.

CONCLUSIONS

We demonstrated the application of AI to assess the abstract and subjective concepts of echocardiographic image quality. The objective evaluation of image quality by AI was also helpful for the intermodality consistency of strain measurements. CC was associated with the precision of automated LVGLS and could potentially be used as a filter to select comparable videos from sequential echocardiographic studies during the whole treatment course of an individual patient (e.g., in the scenario of detecting CTRCD).

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work was partially supported by the Ministry of Science and Technology of Taiwan (grant MOST 107-2314-B-002-257-MY3). The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr. Lung-Chun Lin, Section of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, No. 7, Chung-Shan S. Road, Taipei 10002, Taiwan. E-mail: anniejou@ms28.hinet.net. Twitter: @William04736744.

PERSPECTIVES

COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS: The CC of computer vision can serve as an echocardiographic quality marker. It was associated with the traditional endocardial border index, and echocardiograms with higher CC values had a higher agreement with CMR -derived strain analysis. The CC could facilitate the automation and democratization of the strain analysis–based detection of CTRCD.

COMPETENCY IN PRACTICE-BASED LEARNING: The CC could provide objective feedback about image quality for novice echocardiographic observers, which could shorten the learning curve, especially in the advanced echocardiographic analysis.

TRANSLATIONAL OUTLOOK: Introducing quality control in combination with the currently available automated quantification analysis software can open the era of AI -assisted echocardiographic reporting, improve workflow efficiency, and facilitate communication between ECLs worldwide.

REFERENCES

1. Kenigsberg B, Wellstein A, Barac A. Left ventricular dysfunction in cancer treatment. J Am Coll Cardiol HF 2018;6:87–95.

2. Dang CT, Yu AF, Jones LW, et al. Cardiac surveillance guidelines for trastuzumab-containing therapy in early-stage breast cancer: getting to the heart of the Matter. J Clin Oncol 2016;34:1030–3.

3. Thavendiranathan P, Poulin F, Lim K-D, Plana JC, Woo A, Marwick TH. Use of myocardial strain imaging by echocardiography for the early detection of cardiotoxicity in patients during and after cancer chemotherapy: a systematic review. J Am Coll Cardiol 2014;63:2751–68.

4. Narayan HK, Finkelman B, French B, et al. Detailed echocardiographic phenotyping in breast cancer patients: associations with ejection fraction decline, recovery, and heart failure symptoms over 3 years of follow-up. Circulation 2017;135:1397–412.

5. Negishi T, Thavendiranathan P, Negishi K, et al. Rationale and design of the strain surveillance of chemotherapy for improving cardiovascular outcomes. J Am Coll Cardiol Img 2018;11:1098–105.

6. Negishi K, Lucas S, Negishi T, Hamilton J, Marwick TH. What is the primary source of discordance in strain measurement between vendors: imaging or analysis? Ultrasound Med Biol 2013;39:714–20.

7. Nagata Y, Kado Y, Onoue T, et al. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. Echo Res Pract 2018;5:27–39.

8. Shiino K, Yamada A, Ischenko M, et al. Intervendor consistency and reproducibility of left ventricular 2D global and regional strain with 2 different high-end ultrasound systems. Eur Heart J Cardiovasc Imaging 2017;18:707–16.

9. Chan J, Shiino K, Obonyo NG, et al. Left ventricular global strain analysis by two-dimensional speckle-tracking echocardiography: the learning curve. J Am Soc Echocardiogr 2017;30:1081–90.

10. Khouri MG, Ky B, Dunn G, et al. Echocardiography core laboratory reproducibility of cardiac safety assessments in cardio-oncology. J Am Soc Echocardiogr 2018;31:361–71.e3.

11. Crowley AL, Yow E, Barnhart HX, et al. Critical review of current approaches for echocardiographic reproducibility and reliability assessment in clinical research. J Am Soc Echocardiogr 2016;29:1144–54.e7.

12. Morbach C, Gelbrich G, Breunig M, et al. Impact of acquisition and interpretation on total inter-observer variability in echocardiography: results from the quality assurance program of the STAAB cohort study. Int J Cardiovasc Imaging 2018;34:1057–65.

13. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation 2018;138:1623–35.

14. Abdi AH, Luong C, Tsang T, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. IEEE Trans Med Imaging 2017;36:1221–30.

15. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2016. Available at: https://arxiv.org/abs/1608.06993. Accessed January 28, 2018.

16. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. Available at: https://arxiv.org/abs/1512.04150. Accessed December 14, 2015.

17. von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. N Engl J Med 2017;377:122–31.

18. Gligorov J, Ataseven B, Verrill M, et al. Safety and tolerability of subcutaneous trastuzumab for the adjuvant treatment of human epidermal growth factor receptor 2-positive early breast cancer: SafeHer phase III study's primary analysis of 2573 patients. Eur J Cancer 2017;82:237–46.

19. Hurvitz SA, Martin M, Symmans WF, et al. Neoadjuvant trastuzumab, pertuzumab, and chemotherapy versus trastuzumab emtansine plus pertuzumab in patients with HER2-positive breast cancer (KRISTINE): a randomised, open-label, multicentre, phase 3 trial. Lancet Oncol 2018;19:115–26.

20. Vo HQ, Marwick TH, Negishi K. MRI-derived myocardial strain measures in normal subjects. J Am Coll Cardiol Img 2018;11:196–205.

21. Kasprzak JD, Paelinck B, Ten Cate FJ, et al. Comparison of native and contrast-enhanced harmonic echocardiography for visualization of left ventricular endocardial border. Am J Cardiol 1999;83:211–7.

22. Yoshitani H, Takeuchi M, Hirose M, et al. Head-to-head comparison of fundamental, tissue harmonic and contrast harmonic imaging with or without an air-filled contrast agent, levovist, for endocardial border delineation in patients with poor quality images. Circ J 2002;66:494–8.

23. Thavendiranathan P, Negishi T, Coté M-A, et al. Single versus standard multiview assessment of global longitudinal strain for the diagnosis of cardiotoxicity during cancer therapy. J Am Coll Cardiol Img 2018;11:1109–18.

24. Nagata Y, Takeuchi M, Mizukoshi K, et al. Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software. J Am Soc Echocardiogr 2015;28:630–41.

25. Zamorano JL, Lancellotti P, Rodriguez Muñoz D, et al. 2016 ESC position paper on cancer treatments and cardiovascular toxicity developed under the auspices of the ESC Committee for Practice Guidelines: The Task Force for cancer treatments and cardiovascular toxicity of the European Society of Cardiology (ESC). Eur Heart J 2016;37:2768–801.

26. Plana JC, Galderisi M, Barac A, et al. Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy: a report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2014;27:911–39.

27. Genovese D, Rashedi N, Weinert L, et al. Machine learning-based three-dimensional echocardiographic quantification of right ventricular size and function: validation against cardiac magnetic resonance. J Am Soc Echocardiogr 2019;32:969–77.

28. Yang JX, Quinn G, Mechanic ON, et al. The impact of IAC-Echo accreditation and required quality assurance initiatives on transthoracic echocardiogram interpretation errors. J Am Coll Cardiol Img 2019;12:2090–2.

29. Thavendiranathan P, Grant AD, Negishi T, Plana JC, Popović ZB, Marwick TH. Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes. J Am Coll Cardiol 2013;61:77–84.

30. Marwick TH. Ejection fraction pros and cons: JACC state-of-the-art review. J Am Coll Cardiol 2018;72:2360–79.

31. Medvedofsky D, Mor-Avi V, Byku I, et al. Three-dimensional echocardiographic automated quantification of left heart chamber volumes using an adaptive analytics algorithm: feasibility and impact of image quality in nonselected patients. J Am Soc Echocardiogr 2017;30:879–85.

32. Kusunose K, Abe T, Haga A, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. J Am Coll Cardiol Img 2020;13:374–81.

KEY WORDS artificial intelligence, automated strain analysis, cancer therapeutics–related cardiac dysfunction, left ventricular global longitudinal strain

APPENDIX For an expanded Methods section, videos, and a supplemental table and figures, please see the online version of this paper.