JACC: CARDIOVASCULAR IMAGING VOL. 14, NO. 10, 2021

MINI-FOCUS: AI, MACHINE LEARNING, AND ECHOCARDIOGRAPHY

Artificial Intelligence for Automatic Measurement of Left Ventricular Strain in Echocardiography

Ivar M. Salte, MD,[a,b] Andreas Østvik, MSc,[c] Erik Smistad, MSc, PhD,[c] Daniela Melichova, MD,[b,d] Thuy Mi Nguyen, MD,[a,b] Sigve Karlsen, MD,[d] Harald Brunvand, MD, PhD,[d] Kristina H. Haugaa, MD, PhD,[b,e] Thor Edvardsen, MD, PhD,[b,e] Lasse Lovstakken, MSc, PhD,[c] Bjørnar Grenne, MD, PhD[c,f]

ABSTRACT

OBJECTIVES This study sought to examine if fully automated measurements of global longitudinal strain (GLS) using a novel motion estimation technology based on deep learning and artificial intelligence (AI) are feasible and comparable with a conventional speckle-tracking application.

BACKGROUND GLS is an important parameter when evaluating left ventricular function. However, analyses of GLS are time consuming and demand expertise, and thus are underused in clinical practice.

METHODS In this study, 200 patients with a wide range of left ventricle (LV) function were included. Three standard apical cine-loops were analyzed using the AI pipeline. The AI method measured GLS and was compared with a commercially available semiautomatic speckle-tracking software (EchoPAC v202, GE Healthcare.

RESULTS The AI method succeeded to both correctly classify all 3 standard apical views and perform timing of cardiac events in 89% of patients. Furthermore, the method successfully performed automatic segmentation, motion estimates, and measurements of GLS in all examinations, across different cardiac pathologies and throughout the spectrum of LV function. GLS was $-12.0 \pm 4.1\%$ for the AI method and $-13.5 \pm 5.3\%$ for the reference method. Bias was $-1.4 \pm 0.3\%$ (95% limits of agreement: 2.3 to -5.1), which is comparable with intervendor studies. The AI method eliminated measurement variability and a complete GLS analysis was processed within 15 s.

CONCLUSIONS Through the range of LV function this novel AI method succeeds, without any operator input, to automatically identify the 3 standard apical views, perform timing of cardiac events, trace the myocardium, perform motion estimation, and measure GLS. Fully automated measurements based on AI could facilitate the clinical implementation of GLS. (J Am Coll Cardiol Img 2021;14:1918–1928) © 2021 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]Department of Medicine, Hospital of Southern Norway, Kristiansand, Norway; [b]Faculty of Medicine, University of Oslo, Oslo, Norway; [c]Centre for Innovative Ultrasound Solutions and Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway; [d]Department of Medicine, Hospital of Southern Norway, Arendal, Norway; [e]Department of Cardiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway; and the [f]Clinic of Cardiology, St. Olavs Hospital, Trondheim, Norway.
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 January 20, 2021; revised manuscript received March 26, 2021, accepted April 15, 2021.

ISSN 1936-878X https://doi.org/10.1016/j.jcmg.2021.04.018

Assessment of left ventricular (LV) function is fundamental for diagnosis, risk stratification, and guidance of treatment in patients with cardiac disease. The European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE) recommend global longitudinal strain (GLS) as a supplement to left ventricular ejection fraction (LVEF) when evaluating LV function (1). As a result, vendors have developed software enabling measurement of GLS (2). Although these methods are semiautomatic, analyses are still time consuming and demand expertise, and thus are underused in everyday clinical practice.

Deep learning, the most recent advancement in artificial intelligence (AI), now enables computers to learn from annotated images and perform fully automated image analysis without any operator input (3). Previous AI and machine learning techniques required explicitly designed pattern recognition features to be created by the designers of the AI, whereas the novel deep learning techniques enable the AI to independently learn the patterns and combinations of patterns in the dataset needed to make accurate predictions, thereby allowing for fully automated calculations previously only possible with extensive manual work. This has caused deep learning neural networks to become the most successful and state of the art method of current AI research (4). Deep learning neural networks has successfully been adapted to perform several specific tasks in echocardiographic image analysis that previously would have needed human input, such as view classification (5,6), timing of events (7), and image segmentation (8). Although AI and deep learning in echocardiography are still in its infancy, there is at present commercially available software solutions that have implemented neural networks for tasks such as view classification and segmentation to provide automatic measurement of GLS. However, the core task of strain imaging, namely motion estimation, is still performed by traditional speckle tracking algorithms. We have recently demonstrated that a deep learning neural network could also be trained to estimate motion in two-dimensional (2D)-echocardiography, and that such a network could be implemented in an end-to-end deep learning AI pipeline for automatic measurements of GLS (9). Compared with traditional speckle tracking, far more sophisticated motion estimation algorithms can be constructed by using deep learning. A deep learning-based motion estimation network could learn to integrate information about different moving speckle patterns and global features of an image and independently learn to differentiate artifacts from true motion. Fully automated GLS measurements based on deep learning for motion estimation have the potential to both reduce time spent on manual tracing and improve reproducibility, and, due to the processing speed of optimized deep learning algorithms, this could eventually enable on-screen measurements in real-time while the operator acquires images. Thus, the field of deep learning represents a paradigm shift in medical imaging and could change how we perform clinical measurements in cardiology.

We hypothesized that a fully automated AI method based on deep learning could, without any operator input, identify and classify the 3 standard apical views, perform event timing, trace the myocardium, perform motion estimation, and calculate GLS, producing comparable results to a commercially available semiautomated speckle-tracking method. The aim of this study was to test this hypothesis in echocardiographic examinations from patients with a wide range of LV function, different cardiac pathologies, and varying image quality.

ABBREVIATIONS AND ACRONYMS

2D = two-dimensional
AI = artificial intelligence
ANN = artificial neural network
ASE = American Society of Echocardiography
B-A = Bland-Altman
EACVI = European Association of Cardiovascular Imaging
ECG = electrocardiogram
ED = end diastole
GLS = global longitudinal strain
LOA = limits of agreement
LV = left ventricle
LVEF = left ventricular ejection fraction
ROI = region of interest

METHODS

STUDY DESIGN. A measurement system comparison study was performed by analysis of 200 echocardiographic examinations. Each examination represented a test for each method, resulting in 2 paired GLS measurements for each examination. The first measurement system consisted of a single experienced observer using a commercially available semiautomatic method for GLS measurements. The second measurement system was a novel AI method measuring GLS without any observer input. A single heart cycle was chosen from each view and the exact same recording and cardiac cycle was used for both methods. Analyses were performed without knowledge of clinical data or previous measurement results.

To assess if agreement between methods was affected by LV function, subgroup analyses were performed by categorizing the 200 between method differences by LV function measured using LVEF (normal LVEF >50%, mildly reduced LVEF 40% to 59%, moderately reduced LVEF 30% to 39%, and severely reduced LVEF <30%). Finally, subgroups were evaluated according to image quality (good, fair, or poor).

The proposed deep learning AI pipeline automatically estimates end diastole (ED) and end systole using a deep learning AI timing network. To explore how this automatic event timing affected the GLS measurements, all examinations were analyzed twice by the AI pipeline. First, the deep learning AI pipeline was performed as proposed including automatic event timing using the AI timing network, Second, analyses using the same AI pipeline was repeated using the event timing defined by the reference method.

Intra- and interobserver reproducibility was assessed in a random subset of 25 patients to illustrate the variability observed when measuring conventional LVEF and GLS by the reference method as compared with the novel AI method. Intraobserver reanalyses of these examinations were performed by the same observer 4 weeks after the initial measurements. An experienced second observer at a different hospital analyzed the same examinations to assess interobserver variability. All reanalyses were performed using the exact same heart cycle and blinded to previous measurements and clinical data.

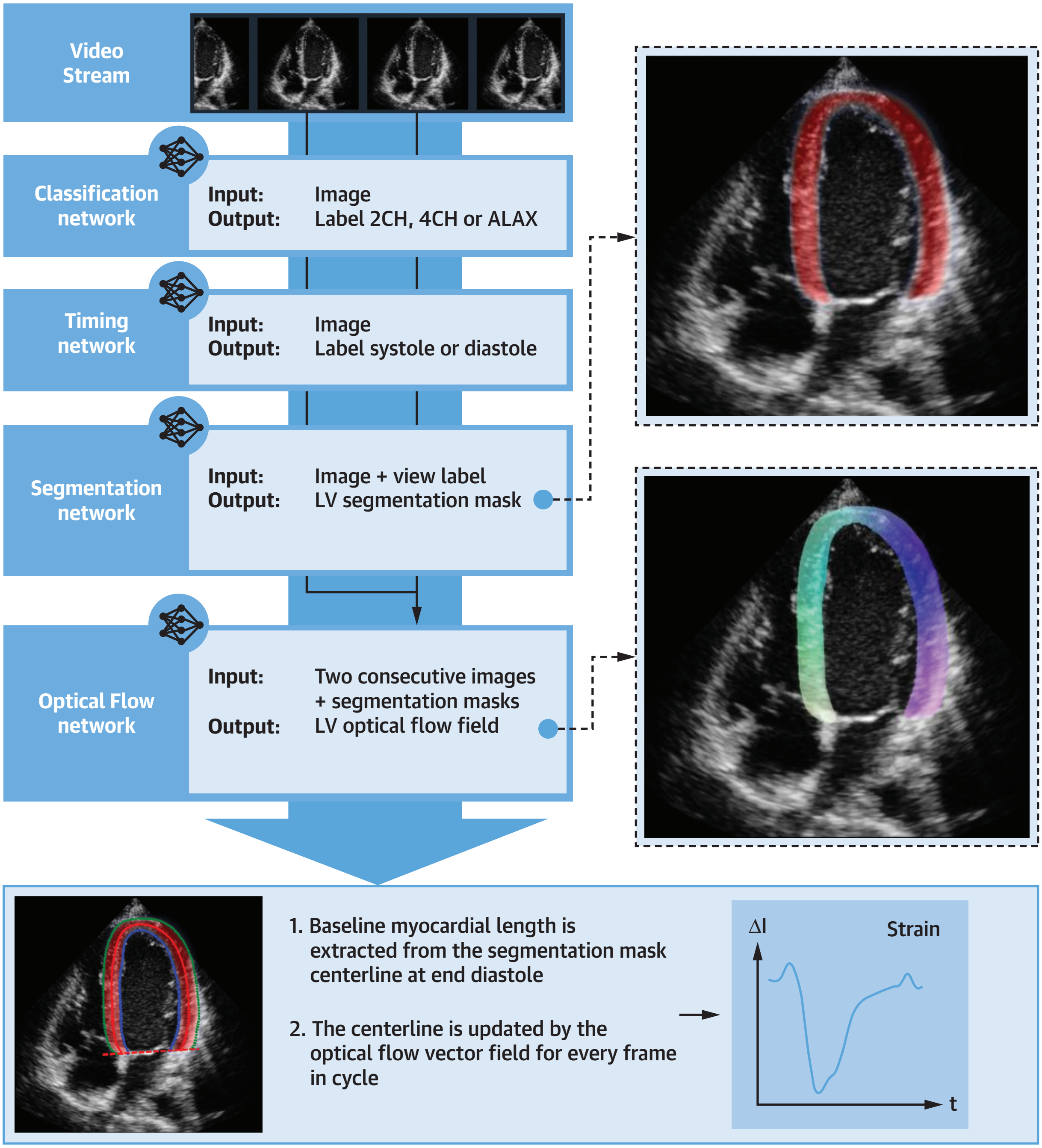
MATERIAL. To achieve a study population with a wide range of cardiac function and different pathologies, we included 5 pre-defined patient groups: 35 patients with non-ST-segment elevation myocardial infarction, 35 patients with ST-segment elevation myocardial infarction, 50 patients with ischemic heart failure, 50 patients with nonischemic heart failure, and 30 patients admitted for chest pain where neither clinical examination, laboratory tests, electrocardiogram (ECG), echocardiography, or coronary angiography revealed any evidence of cardiac origin. Patients were included regardless of the quality of echocardiographic recordings. Myocardial infarction was defined according to the universal definition (10). Patients were included consecutively for each group and regardless of image quality. Exclusion criteria were significant valvular disease, atrial fibrillation, age younger than 18 years, or inability to give written informed consent. The study was approved by the Regional Committee for Medical and Health Research Ethics and was conducted in compliance with the ethical principles of the Declaration of Helsinki.

ECHOCARDIOGRAPHIC EXAMINATIONS. The echocardiographic examinations were recorded using GE Vivid E7/E9/E95 ultrasound systems (GE Ultrasound, Horten, Norway). Echocardiographic examinations and measurements were performed in accordance with EACVI guidelines (11). LV-focused echocardiographic recordings were performed in the 3 standard apical views with simultaneous ECG tracing. Frame rate was 67 ± 9 frames/s. LVEF was measured using the Simpson biplane disc summation method using tracings from apical 4-chamber and 2-chamber views. Image quality was assessed based on visual assessment of each of the 18 individual myocardial segments of the 3 apical views. A segment was considered missing if partly outside the image sector or if the myocardium was indistinguishable from surrounding structures due to artifacts. Good quality examination was defined as no missing segments in either of the 3 apical views, fair quality was defined as 1 to 2 missing segments, and poor quality was defined as >2 missing segments.

STRAIN MEASUREMENTS USING THE REFERENCE METHOD. Conventional GLS was measured using speckle-tracking analyses using the semiautomatic analysis method (2DS) implemented in widely used commercially available software (EchoPAC SWO version 202, GE Ultrasound). ED was defined by the automatic ECG trigger algorithm of the analysis software and only corrected if the automatic QRS detection failed. End-systole was manually defined by the aortic valve closure signal obtained by pulsed-wave Doppler in the left ventricular outflow tract or from continuous-wave Doppler through the aortic valve. The observer manually corrected the region of interest (ROI) by visual assessment of the endocardial and epicardial borders. Spatial and temporal smoothing were kept at default values. Drift compensation was applied as by default. No segments were excluded. A single heart cycle was analyzed for each of the 3 standard apical views and peak strain was obtained as calculated by the software. GLS was calculated as the average peak strain of the 3 apical views. The speckle-tracking analyses were performed in accordance with the consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging (12).

STRAIN MEASUREMENTS USING A DEEP LEARNING AI PIPELINE. We used an in-house-developed AI method based on deep learning consisting of a pipeline of 4 artificial neural networks (ANN) (**Central Illustration**). A detailed technological description of the pipeline has been published separately (9). The first network was based on the Inception and DenseNet architectures and used for image classification. This network was trained to classify a presented image into one of the multiple-view classes, including: 2-chamber, 4-chamber, and apical long axis. The second network was based on a recurrent ANN architecture and was used for event timing. This network was trained to classify series of presented images into systole or diastole. The third network was based on the U-net architecture and used for image segmentation. This network was trained to classify the image per pixel into 4 segmentation classes: lumen, LV myocardium, left atrium, or other/background. Per pixel

CENTRAL ILLUSTRATION Automated Deep Learning Artificial Intelligence Pipeline



Salte, I.M. et al. J Am Coll Cardiol Img. 2021;14(10):1918-1928.

The artificial intelligence (AI) pipeline for automatic measurement of global longitudinal strain consisting of 4 artificial neural networks. Visual feedback of the key steps involved in calculating the global longitudinal strain (GLS) is illustrated, such as the segmentation used to initiate the region of interest, an optical flow field visualizing the predicted local velocities, the extracted centerline from the segmentation mask, and the points visualizing the motion used to calculate GLS; LV = left ventricle.

JACC: CARDIOVASCULAR IMAGING, VOL. 14, NO. 10, 2021
OCTOBER 2021:1918-1928

predictions were used to extract the position, size, and shape of the ventricular myocardium, lumen, and left atrium in an image. The myocardial segmentation performed by this network had a previously reported Dice Coefficient of 0.79 ± 0.08 and was used to initialize the ROI. The fourth network was based on a modified Pyramidal processing, Warping and Cost volume Network (PWC-Net) optimized for estimation of motion in echocardiographic images. This network learned to find patterns in 2 consecutive images and was trained to output an optical flow vector field of equal size as the input images that, when applied to the patterns in the first image, would best reconstruct the location of the same patterns in the second image.

The view classification network was trained on an in-house dataset of out-patient examinations containing 424 hand-labeled echocardiographic recordings. The timing and segmentation networks were trained using the publicly available Cardiac Acquisitions for Multi-structure Ultrasound Segmentation dataset of 500 hand-labeled echocardiographic recordings (13). The optical flow network was trained using synthetic echocardiography images where the true motion was known (14).

The AI method measured strain frame by frame based on the estimated movement of equally spaced points initialized along the centerline from myocardial segmentation at ED. The tracking was performed by updating the position of these points using the displacement fields from the optical flow network. A spline was fitted to the centerline points for each frame. The GLS was calculated for each view as the percentage change in length of the spline from ED to its shortest length through cycle. Similar to the reference method, lagrangian peak strain was calculated for the specified heart cycle in each of the 3 apical views and GLS was calculated as the average of these 3 values.

STATISTICS. Association between methods was estimated by calculating the Pearson correlation coefficient. The mean absolute difference between the 2 measurement systems was calculated using the mean value of the absolute difference between all measurement pairs. The agreement of the paired measurements was assessed using a Bland-Altman (B-A) analysis, which is the recommended statistical method in measurement comparison studies (15). An a priori maximum limit of agreement (LOA) of ±4% was chosen based on known intervendor variability (2). A sample size of 200 subjects was chosen; this sample size provides sufficient accuracy, with 95% confidence interval about the LOA of approximately ±0.24 SDs. Tests for normality were performed using Shapiro-Wilk and Kolmogorov-Smirnov tests. Brown-Forsythe test was used to assess if there was a statistically significant difference in variance between subgroups of measurement pairs when categorized using LVEF and image quality.

TABLE 1 Study Population (N = 200)

Study cohorts	
NSTEMI	35 (17.5)
STEMI	35 (17.5)
Ischemic heart failure	50 (25)
Nonischemic heart failure	50 (25)
No significant cardiac disease	30 (15)
Demographics	
Age, yrs	61 ± 14 (22-91)
Male	144 (72)
Clinical characteristics	
BMI, kg/m^2	27 ± 4 (18-43)
Heart rate, beats/min	74 ± 15 (44-132)
SBT, mm Hg	125 ± 21 (86-197)
Echocardiographic measurements	
Echocardiographic LVEF, %	42 ± 13 (7-70)
LVEDV, ml	128 ± 66 (47-372)
LVESV, ml	80 ± 57 (19-306)
LV function by LVEF category	
Severely reduced: <30%	29 (14.5)
Moderately reduced: 30-39%	60 (30)
Mildly reduced: 40-49%	41 (20.5)
Normal: >50%	70 (35)
Image quality	
Poor (>2 segment missing)	39 (19.5)
Fair (1-2 segments missing)	71 (35.5)
Good (0 segments missing)	90 (45)

Values are n (%) or mean ± SD (range).

BMI = body mass index; LVEDV = left ventricular end-diastolic volume; LVEF = left ventricular ejection fraction; LVESV = left ventricular end-systolic volume; NSTEMI = non-ST-segment elevation myocardial infarction; SBT = systolic blood pressure; STEMI = ST-segment elevation myocardial infarction.

B-A statistical calculations and plot were performed using Python 3.7.4 (Python Software Foundation), where exact 95% confidence interval limits of the LOA were calculated using code based on the method proposed by Shieh (16). All other statistical analyses were performed using open-source statistical Python packages (SciPy 1.5.4 and Statsmodels 0.12.1).

RESULTS

Patient characteristics are summarized in **Table 1**. The view classification network succeeded to classify the correct view in 97% of the cine-loops (584/600). A confusion matrix summarizing classification results for each view is presented in Supplemental Table 1. The timing network succeeded in estimation of both ED and end-systole in 98% (593 of 600) of the cine-loops. Difference in timing of end systole and ED between the deep learning AI timing network and the reference method was 1.8 ± 2.7 frames (17 ± 42 ms) and 0.5 ± 2.7 frames, respectively. Detailed results of the timing network for each view are presented in Supplemental Table 2. When running the AI pipeline as proposed, with event timing defined by the deep learning AI network, the mean difference in measured GLS compared with using event timing defined by the reference method was 0.3 ± 0.3% ($p = 0.02$). A B-A analysis presenting impact of timing method is presented in Supplemental Figure 1. Twenty-one patients (11%) had failures in either the view classification or timing. Both the reference method and the AI method succeeded in measuring GLS in all included recordings, when correct view and timing were verified.

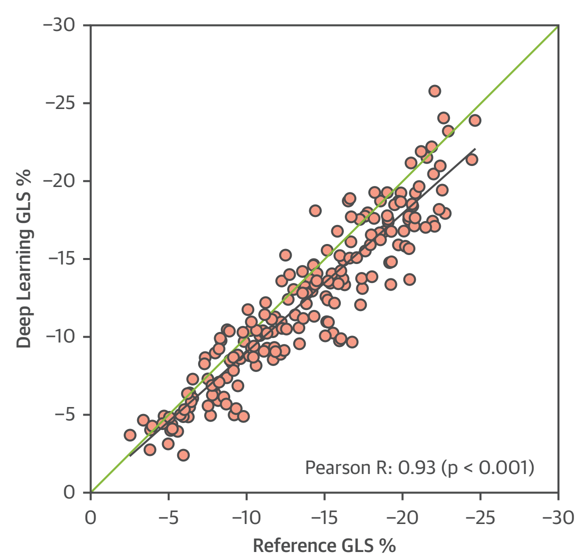
The proposed method was run on a standard desktop computer with a modern graphics card and used approximately 4 ms per frame for view classification, 16 ms per frame for event timing, 10 ms per frame for myocardial segmentation, and 30 ms per frame for motion estimation. Total processing time when running the entire pipeline was 4.3 ± 0.7 s per view and 13.0 ± 2.0 s for a full patient analysis including all 3 apical views.

IN BETWEEN METHODS AGREEMENT. Mean GLS in the entire population was −12.1 ± 5.0% and −13.5 ± 5.3% for the AI method and the conventional method, respectively. The median absolute deviation was 1.4% and mean absolute difference was 1.8 ± 1.5%. There was a highly significant correlation between the methods (Pearson coefficient 0.93; $p < 0.01$, **Figure 1**). The B-A analysis of between method differences revealed a bias of −1.4 ± 0.3% ($p < 0.01$) with estimated LOA of ±3.7% (**Figure 2**).

AGREEMENT CATEGORIZED USING LVEF AND IMAGE QUALITY. The spread of subjects across different categories of LVEF and image quality is presented in **Table 1**. There was no significant difference in variance between measurement pairs from different subgroups categorized using LVEF ($p = 0.06$). Moreover, no significant difference in variance was found between subgroups when categorized using image quality ($p = 0.58$). **Figure 3** presents the B-A plot where measurement pairs are categorically labeled using LVEF and image quality, illustrating the distribution of these categories throughout the range of GLS measured in the study population.

INTRA- AND INTEROBSERVER VARIABILITY AND AGREEMENT. **Figure 4** shows B-A plots illustrating the relative difference in repeated GLS measurements when measured by 2 observers using the reference method, one observer using the reference method and repeated measurements by the automated deep learning AI pipeline. Importantly, the AI pipeline had no operator input and deep learning algorithms are deterministic in design, thus there were no variability when reanalyzing the exact same images. Assessment of intraobserver variability using the reference method resulted in no significant bias −0.1 ± 0.2 ($p = 0.55$) and LOA ±0.5%. However, there was a small interobserver bias observed when using the reference method 0.5 ± 0.4 ($p = 0.04$) and LOA ±2.1%. Visual representations of measurement agreement using the reference method are presented in B-A plots in Supplemental Figures 2 and 3.

FIGURE 1 Correlation Plot

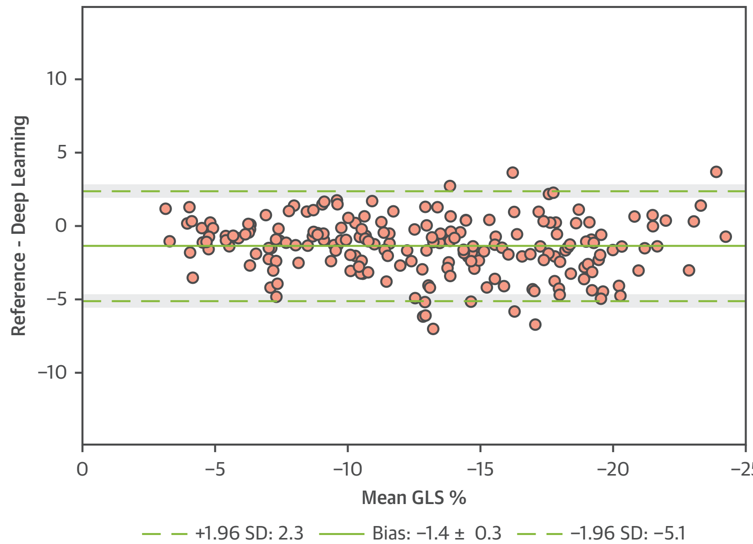


Two hundred echocardiographic examinations measured using both the reference method and the novel artificial intelligence (AI) method. Each **marker** represents one examination. **Solid black line** represents the best fit line to the data by linear regression. **Solid green line** represents the theoretical perfect correlation. GLS = global longitudinal strain.

DISCUSSION

The current study presents, for the first time, the clinical feasibility of an end-to-end AI pipeline that incorporates a deep learning based ANN specifically trained for motion estimation as an alternative to traditional speckle-tracking-based measures of strain. Through a wide range of LV function and image quality, the AI pipeline succeeded without any human input to correctly classify cardiac views and perform timing of cardiac events, and it was able to

FIGURE 2 **Measurement Comparison Results**



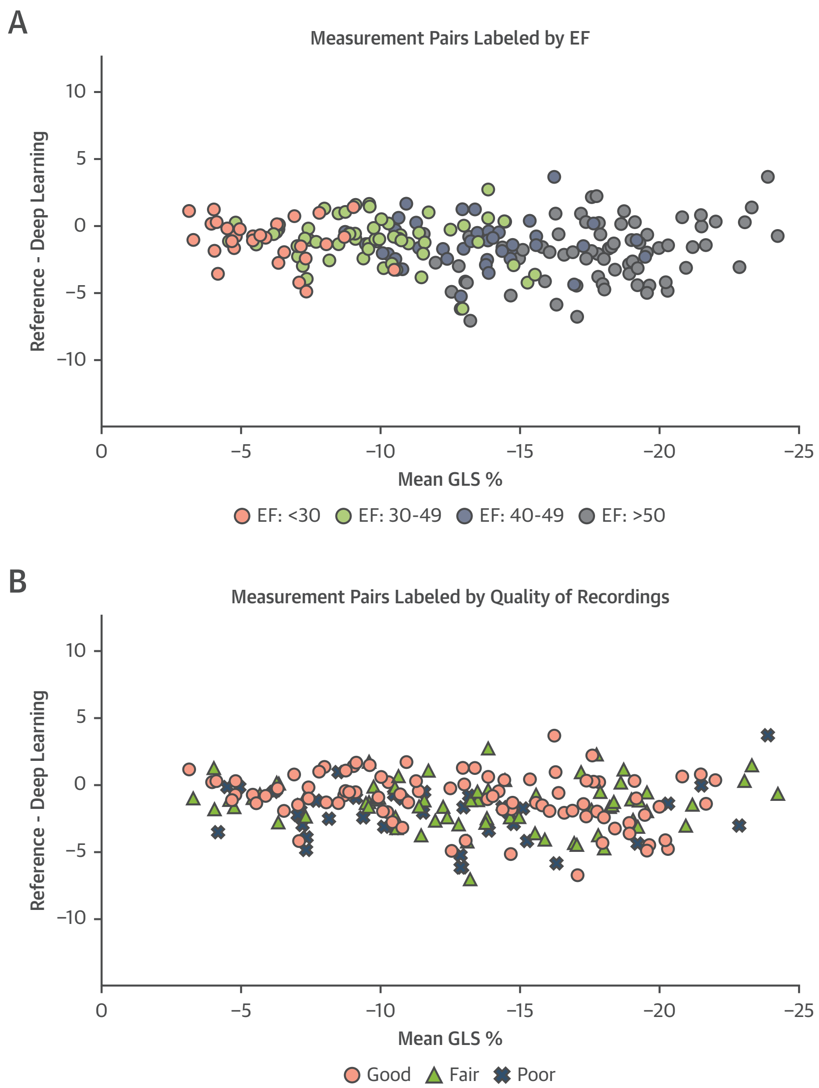
Bland-Altman plot presenting measurement comparison results using 200 echocardiographic examinations comparing the reference method and the AI method. The figure show the limits of agreement (LOA) calculated assuming normal distribution of the differences between methods. Abbreviations as in **Figure 1**.

trace myocardium, estimate motion, and ultimately measure GLS.

The main motivation for developing an AI-based pipeline for GLS measurements is to provide a more robust and automated method, with the potential to provide fully automatic real-time GLS measurements while performing the image acquisition, and with improved tracking accuracy and reduced measurement variability. The currently most widely used semiautomatic speckle-tracking methods need several steps of operator input, and time spent performing a single GLS analysis is reported to range from 5 to 10 min (17,18). In contrast, all steps in the AI pipeline were performed in <15 s. The novel deep learning AI pipeline eliminates the need for time-consuming manual input, which makes it effortless to acquire average measurements from multiple cardiac cycles as recommended in guidelines. Moreover, deep learning algorithms are deterministic. This means that the same input images always gives identical output, without variability (**Figure 4**).

It is important to emphasize that removing the interpretation variability completely by a deterministic deep learning algorithm does make repeated measurements more reproducible but does not necessarily make the measurements more accurate. A measurement error made by a deep learning method will be reproduced every time the method is reapplied to analyze the exact same image. Poor image quality in echocardiography is a common problem and in the present study a total of 19% of subjects had more than 2 of 18 segments missing. The high percentage of examinations with suboptimal image quality could explain that 11% of subjects had failures of either the view classification network or the timing network. Suboptimal image quality is an unavoidable factor that limits the achievable accuracy of both manual and automatic measurements. Thus, measurement error or failure of deep learning algorithms is inevitable, even if deep learning algorithms were to outperform humans in terms of precision of measurements. This underlines the importance of not choosing a fully “black-box” AI method, such as directly predicting LVEF or GLS from images equivalent to “eyeballing” as proposed by some authors (19,20). An AI method should be designed to give visual feedback to the observer. Misclassification of view or timing could easily be corrected by the observer. Motion estimation involves complex calculations that are not directly available for the user to inspect, both with the AI method and the semiautomatic reference method. However, an observer could visually inspect if tracking and motion estimates seem reasonable if provided a visual feedback. The method presented in this study was able to give visual feedback for each frame of the left ventricular

FIGURE 3 **Measurement Comparison Results Categorized**



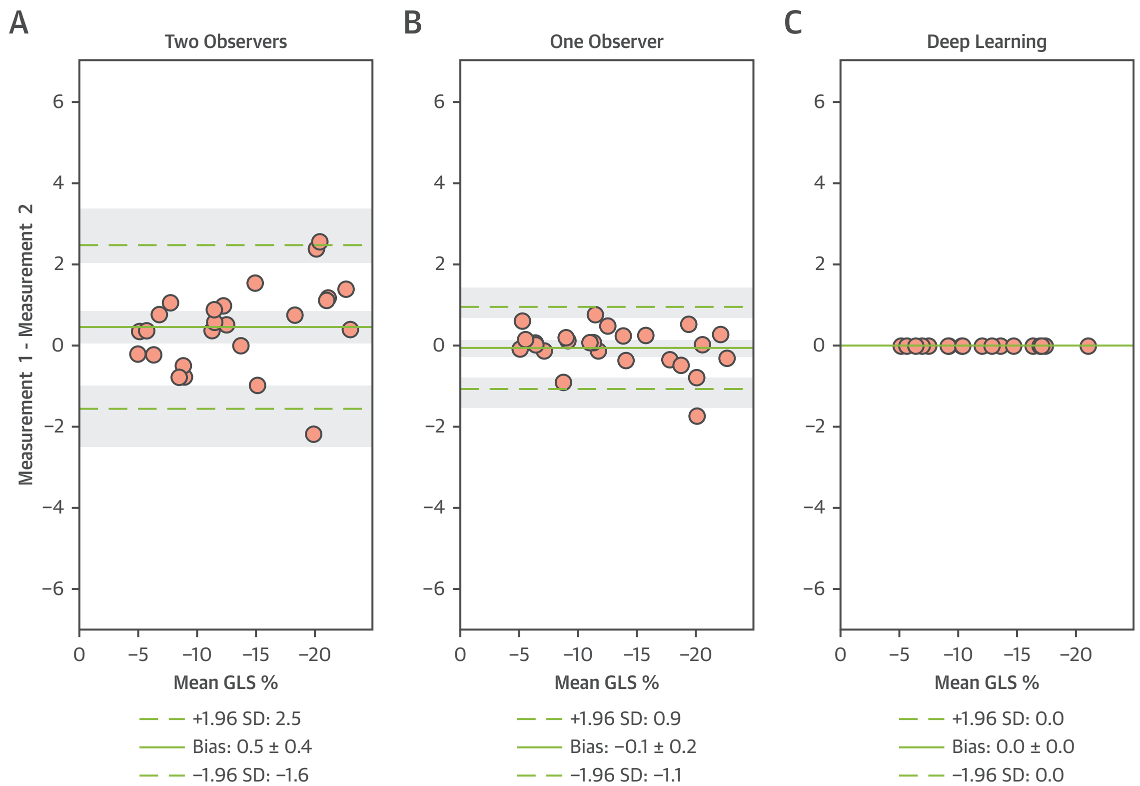
Bland-Altman plot presenting measurement comparison results using 200 echocardiographic examinations comparing the reference method and the novel method based on AI. Measurement pairs labelled by left ventricular ejection fraction (LVEF) **(A)**, and by assessment of image quality **(B)**. GLS = global longitudinal strain.

segmentation, the motion estimation flow field, and the movement of the points used to calculate GLS (**Central Illustration**).

Compared with a previously conducted intervendor comparison study by the EASCVI/ASE/Industry Task Force to standardize deformation imaging (2), the 2 methods in the present study showed excellent correlation and high level of agreement. When assessing agreement, no significant difference in variance was found when measurement differences were categorized according to LV function measured using LVEF and by degree of image quality, suggesting that myocardial dysfunction and image quality had limited effect on agreement.

In the landmark intervendor study by Farsalinos et al. (2), 2 vendor independent software packages were compared with the same reference software as in our study. The bias reported was −0.7% and 2.5% strain units and LOA ±3.4% and ±3.8%, respectively. A study by Nagata et al. (21) compared 2 different vendor independent software packages using the same reference vendor as in our study. They reported

FIGURE 4 Repeated Measurements



Bland-Altman plots of absolute GLS difference in repeated measurements using the exact same video clips in 25 echocardiographic examinations. The figure show agreement between repeated measurements using the reference method by 2 observers **(A)** and 1 observer **(B)**, and for illustrative purposes the expected zero variability of repeated measurements by the fully automated and deterministic AI method **(C)**. The **grey shaded areas** represent the 95% confidence interval (CI) of estimates. Abbreviations as in Figures 1 and 2.

bias of −2.1% and LOA of ±4.1%. Anwar et al. (22) also compared a vendor independent software package to the same reference software as in our study, and found a bias of −2.9% strain units and LOA ±5.5%. Thus, the bias of −1.4% and LOA of ±3.7% observed in our study are well within the overall range of the bias and LOA previously reported in intervendor agreement studies. This provides reasonable evidence to support that GLS measurements by the present deep learning AI pipeline are comparable with other clinically available semiautomatic methods.

To the best of our knowledge, there is currently no clinically available software application for fully automated GLS measurements in 2D echocardiography that have implemented a deep learning neural network specifically to estimate motion and produce a motion flow field as an alternative to traditional speckle-tracking strain algorithms. Except for our technical paper describing the present AI pipeline (9), there are no published journal articles presenting a deep learning neural network for local motion estimation in 2D echocardiography that could produce flow field motion estimates of the entire myocardium. We are only aware of one other journal article that presents an in-detail description of a fully automated AI method for GLS measurements (23). However, although the authors used deep learning to automatically initialize a ROI, conventional optical flow and not deep learning was used for motion estimation and calculation of GLS. Thus, they did not gain the full potential of deep learning to improve measurements of GLS. They found median absolute deviation in GLS measurements of 1.4% in a population of 419 examinations and 1.6% in another population of 110 examinations. These findings are in line with our study where the median absolute deviation was 1.2%. Their method resulted in a GLS processing time of 1 to 4 min per view depending on number of frames and image size, whereas the pipeline in our study used <5 s per view. In addition, the individual networks used in the pipeline succeeded to process frames within milliseconds. Thus, if these deep learning methods are implemented into ultrasound machines, the individual steps of the AI pipeline could be computed during acquisition of images, enabling rapid bedside analysis, and even real-time measurements on the ultrasound scanner.

STUDY LIMITATIONS. We only compared measurements against one reference method. There is no gold standard for GLS measurements and intervendor variability is a known problem. Moreover, the tracking software used by the vendors are not open source. Thus, we cannot conclude whether one measurement system in this study is more accurate than the other. We could only conclude that the measurement variability between these 2 measurement systems is within the range of previous intervendor studies. The statistical power for comparing agreement of measurement across subgroups of LVEF and image quality were limited, due to relatively small sample sizes of subgroup analyses. The total test-retest reliability of an echocardiographic measurement depends on multiple factors related to both image acquisition and observer interpretation. A recent study concluded that acquisition and reader influenced the variability of both GLS and LVEF measurements to a similar extent (24). This suggests that automation of measurements could substantially reduce the total variability in a test-retest setting by removing the individual observer interpretation. We focused on image interpretation and the 2 measurement systems analyzed the exact same images from one predefined cardiac cycle. Hence the present study was not designed to determine the effect of image acquisition on measurement reproducibility. The current deep learning pipeline was not designed to measure multiple cycles. Future work should be done to evaluate implementation of beat-to-beat variation in the algorithm. Another limitation in the present study is that all examinations used for testing the deep learning algorithms were acquired using ultrasound machines from the same vendor. Consequently, we cannot conclude whether the AI method performs equally on images from different vendors. Another topic for further studies is whether deep learning-based strain estimation is more accurate and robust in terms of capturing subtle differences in strain, or when exposed to image artifacts compared with currently used speckle-tracking methods.

Further research is needed to address the mentioned limitations before deep learning measurements could be routinely used in a clinical setting. However, we find the present results promising both in terms of feasibility and agreement with the reference method.

CONCLUSIONS

Fully automated measurements of GLS using a novel deep learning AI-based technology for motion estimation are feasible and fast and they yield results comparable with the most widely used semiautomatic software. Deep learning networks remove the need for manual tracing and could both increase efficiency and improve reproducibility. The system can potentially be implemented in ultrasound scanners and allow for real-time GLS calculations. Fully automated measurements based on AI could be an important step to further facilitate the implementation of GLS in clinical practice.

ACKNOWLEDGMENT The authors thank Håvard Dalen for critical revision of the manuscript.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work was supported by the Research Council of Norway (Project number 237887), Norwegian Health Association, South-Eastern Norway regional health authority, national program for clinical therapy research (project number 2017207), and Centre for Innovative Ultrasound Solutions, a Norwegian Research Council center for research-based innovation. All authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr. Bjørnar Grenne, St. Olavs Hospital, Trondheim University Hospital, Clinic of Cardiology, Postbox 3250 Torgarden, NO-7006 Trondheim, Norway. E-mail: bjornar.grenne@ntnu.no. Twitter: @NTNU, @CIUS_NTNU, @UIO, @UniOslo_MED, @ProCardio1.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: AI methods using deep learning could without any operator input automatically estimate myocardial deformation and measure GLS. AI methods could reduce measurement variability and provide more robust measurements of LV function.

TRANSLATIONAL OUTLOOK: Future research should assess clinical implementation of AI-based methods in ultrasound scanners for fully automatic real-time on-screen measurement of GLS.

REFERENCES

1. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16:233–270.

2. Farsalinos KE, Daraban AM, Unlu S, Thomas JD, Badano LP, Voigt JU. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: the EACVI/ASE Inter-Vendor Comparison Study. *J Am Soc Echocardiogr*. 2015;28:1171–1181, e1172.

3. Litjens G, Ciompi F, Wolterink JM, et al. State-of-the-art deep learning in cardiovascular image analysis. *J Am Coll Cardiol Img*. 2019;12:1549–1565.

4. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.

5. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med*. 2018;1:1–8.

6. Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. *Ultrasound Med Biol*. 2019;45:374–384.

7. Meidell Fiorito A, Østvik A, Smistad E, Leclerc S, Bernard O, Lovstakken L. Detection of cardiac events in echocardiography using 3D convolutional recurrent neural networks. 2018 IEEE International Ultrasonics Symposium (IUS). *Kobe* 2018:1–4.

8. Chen C, Qin C, Qiu H, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. 2020;7:1–33.

9. Østvik A, Salte IM, Smistad E, et al. Myocardial function imaging in echocardiography using deep learning. *IEEE Trans Med Imaging*. 2021;40(5):1340–1351.

10. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. *Eur Heart J*. 2012;33:2551–2567.

11. Galderisi M, Cosyns B, Edvardsen T, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2017;18:1301–1310.

12. Voigt JU, Pedrizzetti G, Lysyansky P, et al. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16:1–11.

13. Leclerc S, Smistad E, Pedrosa J, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging*. 2019;38:2198–2210.

14. Alessandrini M, Chakraborty B, Heyde B, et al. Realistic vendor-specific synthetic ultrasound data for quality assurance of 2-D speckle tracking echocardiography: simulation pipeline and open access database. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2018;65:411–422.

15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.

16. Shieh G. The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement. *BMC Med Res Methodol*. 2018;18:1–11.

17. Barbier P, Mirea O, Cefalu C, Maltagliati A, Savioli G, Guglielmo M. Reliability and feasibility of longitudinal AFI global and segmental strain compared with 2D left ventricular volumes and ejection fraction: intra- and inter-operator, test-retest, and inter-cycle reproducibility. *Eur Heart J Cardiovasc Img*. 2015;16:642–652.

18. Manovel A, Dawson D, Smith B, Nihoyannopoulos P. Assessment of left ventricular function by different speckle-tracking software. *Eur J Echocardiogr*. 2010;11:417–421.

19. Asch FM, Poilvert N, Abraham T, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circ Cardiovasc Img*. 2019;12:e009303.

20. Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. *NPJ Digital Med*. 2020;3:1–10.

21. 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;28:630–641.

22. Anwar S, Negishi K, Borowszki A, et al. Comparison of two-dimensional strain analysis using vendor-independent and vendor-specific software in adult and pediatric patients. *JRSM Cardiovasc Dis*. 2017;6:1–11.

23. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138:1623–1635.

24. Baron T, Berglund L, Hedin EM, Flachskampf FA. Test-retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. *Clin Res Cardiol*. 2019;108:355–365.

KEY WORDS artificial intelligence, artificial neural networks, deep learning, echocardiography, global longitudinal strain, machine learning

APPENDIX For supplemental tables and figures, please see the online version of this paper.