Quality assurance of segmental strain values provided by commercial 2D speckle tracking echocardiography using in-silico models

A Report from the EACVI-ASE Strain Standardization Task Force

Assami Rösner, MD, PhD[1,2]; Martino Alessandrini, PhD[3]; Didrik Kjønås, MD[1]; Oana Mirea, MD, PhD[3,6]; Sandro Queirós, PhD[3,4,5]; Jan D`hooge, PhD[3]

[1]Department of Cardiology, University Hospital North Norway, Tromsø, Norway

[2]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

[3]Department of Cardiovascular Sciences, KU Leuven, Belgium

[4]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[5] ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal

[6] Department of Cardiology, University of Medicine and Pharmacy, Craiova, Romania

Correspondence:

Jan D'hooge PhD

Department of Cardiovascular Sciences – University Hospital Gasthuisberg

Lab on cardiovascular imaging and dynamics – Medical Imaging Research Center (MIRC)

3000 Leuven, Belgium

Tel: +32.16.34.90.12

E-mail: jan.dhooge@uzleuven.be

ABSTRACT

The aim of this study was to determine the accuracy and reproducibility of vendor-specific regional strain values by echocardiography using in-silico data.

Synthetic two-dimensional ultrasound grey-scale images of the left ventricle (LV) were generated, knowing the longitudinal segmental strain values from the underlying electro-mechanical LV-model. Four out of five models mimicked transmural infarctions with systolic segmental stretching in different vascular areas. Cine loops in the three apical views were synthetically generated at four noise levels. All in-silico images were repeatedly analyzed by a single investigator and some by another investigator.

The absolute errors varied significantly between vendors from 3.3±3.1% to 11.2±5.9%. The area under the curve for the identification of segmental stretching ranged from 0.80 (CI 0.77-0.83) to 0.96 (CI 0.95-0.98). The levels of agreement for the intra- investigator variability varied between -3.0 to 2.9% and -5.2 to 4.8%, and for the inter-investigator variability between -3.6 to 3.5% and -14.5 to 8.5%.

Segmental strain analysis allows the identification of areas with segmental stretching with good accuracy. However, single segmental peak-strain values are not accurate and should be interpreted with caution. Nevertheless, our results indicate the usefulness of semi-quantitative strain assessment for the detection of regional dysfunction.

KEY WORDS:

Speckle tracking
Insilico simulated model
2D strain imaging
Quality assessment
Different vendors

INTRODUCTION

Strain imaging by speckle tracking has been introduced more than two decades ago(Leitman, et al. 2004, Amundsen, et al. 2006). Since, several thousand articles have been published about speckle tracking echocardiography (STE), with increasing impact on clinical management of patients in cardiology. In these studies, typical patterns of differing segmental strain curves in the left ventricle have been described for the assessment of acute and chronic ischemic heart disease(Stoylen, et al. 2000, Hanekom, et al. 2005, Miyasaka, et al. 2005, Grenne, et al. 2010, Rosner, et al. 2015), different forms of ventricular hypertrophy(Cikes, et al. 2010), storage diseases(Phelan, et al. 2012), cardiomyopathies(Haugaa, et al. 2012), transplant rejection(Marciniak, et al. 2007), cardiac dyssynchrony(Parsai, et al. 2009) congenital heart disease(Friedberg and Mertens 2009) and risk evaluation for cardiac arrhythmias(Haugaa, et al. 2012) to name a few. At present, two-dimensional (2D) STE is implemented on most ultrasound systems and has become the most frequently used tool for non-invasive assessment of regional myocardial function. STE is based on frame-by frame tracking of echo-dense speckles within the myocardium with subsequent calculation of regional myocardial deformation from the obtained motion field (Leitman, et al. 2004, Amundsen, et al. 2006). Some of the commercial solutions have been validated by comparing STE derived strain values with values obtained by MRI tagging and sonomicrometry(Amundsen, et al. 2006). However, there remains uncertainty about the accuracy of the reference techniques and many of the commercially available STE algorithms have not been validated in this way. Other studies used MRI with late-gadolinium enhancement as the gold standard for the presence of reduced myocardial function(Zhang, et al. 2005, Thorstensen, et al. 2012, Mirea, et al. 2018, 2018). However, this MRI methodology does not assess myocardial deformation directly and thus remains limited to validation of the detection of severe myocardial dysfunctional areas only. Thus, although global strain values were shown to be reliable in both synthetic (D'hooge, et al. 2016) and clinical (Farsalinos, et al. 2015) validation studies, quality assurance of the different STE solutions for the assessment of *regional* strain in a standardized and reliable way is still an unsolved problem.

With this goal in mind, an in-silico model was recently developed (Alessandrini, et al. 2018). This model combines an electromechanical model of the left and right ventricles (Marchesseau, et al. 2013) with an ultra-realistic ultrasound simulation method(Alessandrini, et al. 2018) in order to generate images synthetically, which provides data sets for STE with a solid ground truth with regionally differing strain values. The model was designed to mimic ischemic hearts with areas of longitudinal stretching according to the typical location of transmural infarctions.

This study was initiated as part of the work of the task force on STE standardization by the European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE) in collaboration with industry (Thomas and Badano 2013, Voigt, et al. 2015). The aim of the study was to investigate the accuracy and variability of segmental strain values of all commercial STE solutions using these synthetic data sets and to investigate their ability to correctly detect and localize regions of segmental stretching.

MATERIALS AND METHODS

Synthetic images and corresponding segmental reference strain values were generated with the simulation pipeline as described in a previous publication(Alessandrini, et al. 2018). In brief, a realistic cardiac anatomy was segmented at end-diastole from a high-resolution cardiac MR scan. The obtained 3D mesh was subsequently made dynamic using a clinically validated electro-mechanical (E/M) finite element model (FEM) of the heart (Marchesseau, et al. 2013). By changing parameter settings of the FEM, five different motion patterns were simulated: four models simulating acute myocardial infarctions with regional systolic segmental stretching according to different locations of vascular occlusions, representing proximal and distal left anterior descendent (LAD), right (RCA) and circumflex (CX) coronary artery occlusion; and one additional model simulated a “normal heart”, i.e. without coronary lesions and thus without segmental stretching (cf. Figure 1). From each of the five numerical E/M simulations, three apical ultrasound recordings were simulated corresponding to a two, three and four-chamber view. Vendor-specific ultrasound speckle texture was generated by using real scans from the respective systems as “templates” in the ultrasound simulation process. Moreover, vendor-specific system settings, such as frame rate, were used in the simulation setup. To test

sensitivity to noise, three noisy versions of each sequence were generated by altering the contrast between myocardium and blood pool according to a contrast to noise ratio of 60 (high quality), 40 (mid quality) and 20 (low quality). Overall, 60 simulated sequences were thus generated for each vendor (5 motion models × 3 apical views × 4 noise values). Example images were presented in the previous publication of Alessandrini et al(Alessandrini, et al. 2018).

INDUSTRY PARTNER RECRUITMENT.

All industry partners within the task force were invited to participate in the study by an open letter. In collaboration with seven ultrasound machine manufacturers, the synthetically generated data were converted into (vendor-specific) DICOM image loops as to enable loading them in the commercial STE solutions. Additionally, 2 manufacturers of generic software solutions for speckle tracking analysis participated in the comparison. For these two manufacturers, DICOM loops from GE (Vingmed Ultrasound, Horton, Norway) were used for analysis, to be consistent with prior in-vivo reports on quality assurance of STE by the task force(D'hooge, et al. 2016, Mirea, et al. 2018, 2018).

SPECKLE TRACKING ANALYSES

One experienced observer (A.R.) analyzed all data for all vendors. The observer had a solid background in speckle tracking analyses with experience in using different strain imaging software packages in clinics and numerous research settings. The observer underwent individual training by each vendor for use of the respective software package right before starting the analyses. All manufacturers and software solutions, with respective release versions, are listed in **Table 1**.

The region of interest was automatically defined by the software and manually corrected as required based on visual inspection. In case no automated contouring solutions were available for a given software product, the ROI was defined manually. The ROI was set at the frame suggested by the software, which was in seven out of nine software packages the first frame of the sequence, while for TOMTEC-ARENA 2DCPA (TOMTEC, Unterschleissheim, Germany) the ROI was drawn at end-systole, and for EchoPac (GE Vingmed Ultrasound, Horten, Norway) in mid-systole. Snapshots of the original contouring and segmental definition for the ground truth data were used to provide correct

definition of the endocardial borders and positioning of the segments during analysis, thereby limiting errors due to misalignment between the manually defined ROI and the original LV model for the ground truth data. Definition of LV segments was automatically done by all software solutions. Tracking was a fully automated process where the tracking results were visually assessed by validating whether the tracked contours sufficiently followed the local wall motion throughout the cardiac cycle. If tracking was not approved, the contour could be redrawn at a maximum two times. When three results were not approved, the last out of the three tracking results was included into the database.

The first frame in the DICOM sequences corresponded to the onset of the cardiac cycle, i.e. end diastole. The time-point of aortic valve closure (AVC) was predefined by the model and equally used for all data analyses of all software-types. Time sequences of segmental strain were shown as curves over one cardiac cycle. Excel datasets with strain values over time were extracted from all software packages. End-systolic (ES) longitudinal strain-values (i.e. at the time-point of AVC) were used after assuring that the AVC time-point was congruent to the predefined AVC. All solutions allowed for drift correction, which was enabled in all analyses. Since subendocardial longitudinal strain was the only measure provided by all nine software solutions, only longitudinal subendocardial strain was investigated in this study even though the synthetic data and several software solutions provide myocardial, epicardial longitudinal and transversal strain as well.

REPRODUCIBILITY

For intra-observer variability, after a time period of 3-5 weeks, the same observer (A.R.) reanalyzed the dataset for all models, views and noise-levels twice, while for inter-observer variability a second independent observer (D.K.) blinded to the analyses of the first observer reanalyzed all imaging loops of two models (CX and distal LAD infarction) once.

The five models with four different noise-levels including 3 projections with 6 segments each were thus repeatedly (i.e. 3 times) analyzed, resulting in 270 segmental analyses per noise level, and a total of 1080 repeated computations of longitudinal peak systolic strain values for each software solution.

STATISTICAL ANALYSES

All statistical analyses were performed using IBM SPSS Statistics 25 software ((IBM Corp., Armonk, N.Y., USA). The mean, standard deviation (SD) and confidence interval (CI) of the absolute error in segmental peak strain were calculated per software solution as the difference between the measured strain values and the ground truth values provided by the model. Similarly, for each vendor, the limits of agreement (LOA) with the ground truth were calculated based on a Bland–Altman analysis. For the assessment of the influence of different factors (i.e. repeated analyses, different noise-levels and models) on the accuracy of the segmental strain measurements, for each vendor, linear mixed models were used. ANOVA with Bonferroni post hoc testing was used to identify differences when appropriate. In addition, receiver operating characteristics (ROC) curves with calculation of the area under the curve (AUC) were used per vendor for segmental classification of stretching / non-stretching segments as a marker of severe myocardial dysfunction. Cut-off values for the highest sensitivities and specificities for the detection of stretching segments were calculated.

In order to verify whether the spatial strain distribution across the ventricle – as a possible hallmark for specific diseases – was correctly detected, the normalized cross correlation coefficient between the measured and ground truth end-systolic bulls-eye plots was calculated. A value of 1 would indicate a perfect match of the observed pattern (irrespective of the strain amplitudes) while a value 0 would indicate no correlation in the spatial patterns at all. Differences between vendors in normalized cross-correlation or AUC were tested while correcting for multiple testing using Bonferroni-Holmes. Finally, to assess intra- and inter-observer variability, the levels of agreement based on Bland–Altman plots were calculated.

RESULTS

Figure 2 shows the average ES longitudinal strain of the measured data including all noise-levels for stretching and non-stretching segments. While the median values for both types of segments were markedly different for some vendors (e.g. Toshiba ACP v. 3.2 (Toshiba, Otawara, Japan), TOMTEC-ARENA 2DCPA (TOMTEC, Unterschleissheim, Germany), they were not for others (e.g. Samsung CardiacPack1.00 0615 (Samsung, Seoul, South Korea), Hitachi 2DTT Analysis v7.0a (Hitachi,

Tokyo, Japan). The absolute segmental error (i.e. the absolute difference between the measured strain and the ground truth of the model) calculated for different noise levels is shown in **Figure 3**. Overall, errors varied from 3.3±3.1% (i.e. QLAB aCMQ (Philips, Andover, Massachusetts, USA) (no noise) to 11.2±5.9% ((CardiacPack1.00 0615 (Samsung, Seoul, South Korea), highest noise-level). One vendor (i.e. Samsung) had a significantly higher error compared to all other vendors. For five out of nine vendors (marked with *) the absolute segmental error showed to be noise dependent (i.e. significant Bonferroni post-hoc testing). For these vendors, as expected, the lowest error was observed in the noise free data and increased with the amount of noise.

Figure 4 and **Appendix, Figure 1** depict the agreement of the measurements as expressed by Bland-Altman plots and regression plots, respectively. All software solutions show a general tendency to underestimate the degree of segmental stretching while most also underestimated the degree of segmental shortening. Samsung showed a significant bias of -10%, resulting in a general overestimation of the estimated strain magnitude. Overall, the limits of agreement varied from about 16% (i.e. Toshiba) to about 22% (i.e. Hitachi).

In **Table 2**, the average normalized cross-correlation coefficient is presented for all noise-levels and vendors. This coefficient expresses the ability of the different software solutions to correctly show the spatial distribution of the end-systolic strain values irrespective of their absolute value. Mean values varied between 0.55 (at the highest noise level) and 0.89 (at the lowest noise level), where no significant differences between noise levels were found but differences between vendors were significant (**Table 3; upper part**).

ROC curve analysis was performed, investigating sensitivities and specificities for correct identification of stretching segments on an individual segmental basis. ROC curves, AUC and vendor-specific cut-off values are shown in **Figure 5** and **Table 4**. Areas under the curve (**Table 4**) show for most of the software solutions good or excellent results for the accurate detection of segmental stretching. All cut-off values were negative, between -2.4 and -4.8%, while CardiacPack1.00 0615 (Samsung, Seoul, South Korea) with -13.1% represented an outlier. Again, statistically significant

differences were found between vendors (**Table 3; lower part**) but not between data sets with a different amount of noise.

Finally, **Tables 5 and 6** show LOAs for intra- and inter-observer variability as an indicator for reproducibility for repeated measurements. Intra-observer reproducibility varied between vendors from 5.9% (GE) to 10% (XStrain2D- v5.50, Esaote, Florence, Italy) absolute differences of strain-estimates. As expected, inter-observer agreement was lower and remained below 10% for two vendors only (i.e. Toshiba, GE). Mixed models analyses for the different effects of repeated measurements showed no significant effect of the time point of investigation, demonstrating no learning bias. Moreover, these models showed no significant impact of the observer on the differences between vendors, models and noise levels.

DISCUSSION

The present study confirms earlier reported variability of segmental strain values, which is known to be higher than that of global strain assessment (Aarsaether, et al. 2012, Rosner, et al. 2015). Indeed, global strain measurements need the correct delineation of the myocardium with subsequent correct tracking mainly of the mitral annulus (showing as a very bright landmark) and the apex (Grue, et al. 2018). These structures are relatively easy to detect automatically on average quality 2D grey scale imaging loops and remain fairly constant in appearance throughout the cardiac cycle. In contrast, segmental strain needs correct tracking of regional speckle patterns at a smaller scale, while the patterns can change appearance during the cardiac cycle. Moreover, correct tracking of the mitral annulus can easily be verified visually(Grue, et al. 2018), while visual quality assurance of segmental tracking of longitudinal strain and differences between neighboring segments is much harder. All combined, compared to global strain, correct segmental strain estimation is much more challenging.

In general, the absolute strain error was found to remain below 5% for most noise levels and for all vendors but one (Figure 3). Given that normal segmental values are around 20%, even an absolute strain error margin of about 4% (as obtained by the most accurate software solutions), seems relatively high. This statement is indeed supported by the limits of agreement found during Bland

Altman (BA) analysis (Figure 4) where even for the best performing solutions, the LOA remained about 16%, which is remarkably high. This study thus shows that values of regional ES strain only should be interpreted with extreme caution when applied to diagnostics or therapy guidance.

Interestingly, these BA plots show a clear trend for all vendors to be biased towards 'no deformation', i.e. both the amount of stretching and the amount of shortening are underestimated (in magnitude), resulting in the error in strain being directly related to the strain magnitude. Most likely, this is the result of spatial and/or temporal smoothing of the regional strain data in order to make the solutions more robust to noise(Rosner, et al. 2015). In this context, it should be pointed out that the synthetically generated data had relatively large differences in strain between neighboring segments (cf. Figure 1), which may have negatively biased the accuracies reported in this study. It can be assumed that reading errors correlate positively with the magnitude of segmental differences and might be much smaller when analyzing healthy populations with homogeneous segmental strains.

The influence of noise on the reported accuracies was vendor-dependent, where 4 out of 9 vendors seemed to have effective noise-reduction algorithms while the remaining 5 vendors showed higher errors with increasing noise levels. Although these findings are indicative of the effect of noise on different solutions, it should be pointed out that the noise imposed on the synthetic data was homogenous in space (i.e. white noise) which could be representative for differences in signal-to-noise ratio of the ultrasound systems but is distinct from clutter noise and artifacts typically observed clinically (e.g. reverberations, shadowing, etc.). The current findings should thus be interpreted with caution as they may change in the presence of more realistic noise characteristics. However, most importantly, also for the solutions that remained sensitive to noise, its influence on the obtained accuracy was relatively small compared to the total strain estimation error suggesting that noise is not the main cause of low accuracy of regional strain estimates.

Although the absolute regional strain values showed to be not sufficiently accurate, their relative differences across the ventricle may carry important information. Hereto, the spatial

distribution of the end-systolic strain pattern across the LV model was compared against the ground truth independent of the absolute strain magnitude using the normalized cross-correlation coefficient (Table 2). Cross-correlation-coefficients of ~0.8 and higher for the best software solutions indicate ability for pattern-recognition, even though direct conclusions for the clinical applicability in different pathological settings cannot be made. For cross-correlations, significant differences between vendors were found (Table 4). These differences seemed to be roughly driven by reaching a correlation value of ~0.80, indicating that some vendors seemed to discover the spatial distribution in strain more truthfully than others. Moderate cross-correlation values (0.6-0.8) likely resulted of excessive spatiotemporal smoothing and measurement error as observed in Figure 4. However, as for the absolute strain error, the relative strain differences across the LV were relatively little influenced by noise, as there was no statistically significant difference in the correlation values between noise levels.

As a test for the ability of an algorithm to detect disease based on segmental strain values, the local strain estimates were used to classify segments between 'stretching' and 'shortening', where the former group was considered representative of severe regional dysfunction irrespective of the absolute strain (stretching) magnitude. Overall, all software solutions but one showed good classification results with an AUC above 80%, with two vendors standing out and reaching an AUC above 90%. These findings seem to indicate that, despite their questionable accuracy for single regional values, regional strain remains useful for detecting severe dysfunctional segments. The latter might be useful for less experienced users but also for expert users, especially when post-systolic shortening might be visually mistaken as systolic contraction.

The LOAs for intra-observer variability were found to range from ~6-10% and were – logically – even larger for inter-observer variability (~7-22%). These values are relatively high, and are in line with previous reports on the reproducibility of segmental strain values(Dalen, et al. 2010, Mirea, et al. 2018).

Given that the only difference between repeated measures was the endocardial border definition (and the associated LV segment definition), it can be expected that automatic segmentation

would be beneficial to increase the reproducibility of the segmental strain estimates. In the present study, it seemed that software solutions with automated border detection and robust tracking algorithms allowed less interference by the reader, pointing to markedly reduced inter-and intra-investigator variabilities as well as lower general segmental reading errors.

STUDY LIMITATIONS

The strain values reported in this study are low with shortening segments rather corresponding to clinically hypokinetic segments (i.e. regional strain values up to -16% only). The reason for these low strain magnitudes is in part caused by the lack of atria in the cardiac mechanical model underlying the synthetically generated data sets as the atrial kick accounts for about 5% of LV strain (i.e. pre-ejection stretch)(Zwanenburg, et al. 2005). In this context, the tested speckle tracking solutions had to operate under extra challenging conditions, as the differences between LV segments were smaller than what would be seen clinically. Although this could negatively bias the reported accuracies, the overall conclusions remain in line with previous reports of the speckle tracking standardization task force(Mirea, et al. 2018). For better reflection of pathologies with mildly to moderately reduced shortening (excluding segmental lengthening), new models will be needed with an appropriate range of strain values as the current models only allowed to test for the detection of severe dysfunction only. Similarly, regional ventricular geometry (e.g., wall thickness) of the E/M model and its spatial heterogeneity may not always have been representative for the clinical situation which may equally have biased our findings.

The synthetically generated sequences made use of real recordings as a template for the acoustic model defining the imaging scene(Alessandrini, et al. 2018). As a result, intrinsic differences between vendors in the quality of these template recordings could thus have induced bias between vendors, which could be particularly true as not all vendors used their flagship systems to record these template images (selected from the in-vivo speckle tracking standardization study(Mirea, et al. 2018, 2018). Similarly, although the synthetic data sets look visually very realistic, further improvements to the simulation framework could be made in order to include more realistic (clutter) noise and thereby avoid bias due to improper noise characteristics. However, the main goal of this study was not to

compare software solutions and give them a rank but rather to provide an independent quality assurance system in which each of the commercial software solutions could be verified. Moreover, this system could be used as a benchmark for continued development of the existing solutions, as well as for potentially new solutions entering the market. In this context, it is worth pointing out that the software solutions used in the present study may not be the latest versions available on the market given the continuous upgrades made on these software packages. Overall however, it is reassuring that the quality metrics of the different software solutions reported align with what was previously reported based on studies in-vivo(Mirea, et al. 2018), making a case for the use of in-silico trials for this type of software verification.

Although the accuracy of segmental strain values were tested in this study, important information may lie in the temporal characteristics of the segmental strain curves either by visual readings (Menet, et al. 2017) or via pattern recognition (Tabassian, et al. 2017, Duchateau, et al. 2020). The accuracy of these temporal profiles was not tested in the current study. Similarly, although the spatial distribution of the segmental strain values across the ventricle can provide valuable information, we could only test the accuracy of detecting these patterns technically by reporting the normalized cross-correlation coefficients with the ground truth distribution. How this impacts clinical readings remains to be tested where initial reports have recently become available(Duchateau, et al. 2020).

Foreshortening (Unlu, et al. 2020), missing parts of segments, and reverberations are known to impact (regional) strain values and although the synthetic data generation framework perfectly allows to test this effect, this fell outside the scope of the current study. Although all data sets analyzed in the present study were therefore idealized recordings in terms of field of view and the absence of clutter, shadowing and reverberation artefacts, future studies are needed to address their relative impact by using specifically designed models.

CLINICAL IMPLICATIONS

The current study reconfirms previous reports that segmental longitudinal ES strain values are not sufficiently accurate and robust in order to be used – in isolation – as a guide for clinical decision-making. Notwithstanding, the study also shows that the detection of regional severe dysfunction (i.e. segmental stretching) is accurate, which can be useful in itself for example to differentiate acute coronary occlusions or transmural coronary lesions in patients with acute coronary syndrome and NSTEMI ECG (Grenne, et al. 2010). Similarly, the differentiation of segmental stretching, segmental shortening or the absence of significant segmental deformation can be useful for viability assessment for echocardiography at rest (Rosner, et al. 2015). Segmental stretching can be present at all kinds of myocardial pathology, even in small regions with pathological changes, while the global function is still preserved as found in mild forms of myocarditis (Hsiao, et al. 2013), cardiomyopathies (Cikes, et al. 2010, Weidemann, et al. 2012), transplant-rejection(Marciniak, et al. 2007) or storage-diseases(Cikes, et al. 2010), where the semi-quantitative evaluation of inhomogeneous bulls-eye patterns and the typical localization of pathological segments are sufficient to render information needed for clinical decision making. The distribution of patterns of dysfunction across the LV was shown to be accurate and probably usable in the clinical setting, although variations between vendors were observed. Similarly, machine learning might better capture the complex spatiotemporal features of the segmental strain data set and thereby strengthen its diagnostic power.

CONCLUSION

Single segmental strain values should be interpreted with caution and should not be used as such for clinical decision-making as they are neither accurate nor precise irrespective of the software solution used. Nevertheless, segmental speckle tracking carries value as a semi-quantitative method, where identification and correct localization of regions with pathologic deformation can be identified. This performance of this semi-quantitative approach seems to vary between vendors but requires specific clinical validation studies.

ACKNOWLEDGEMENTS

We thank the Northern Norway Regional Health Authority for supporting the first author with a grant (ID 6884/ SFP1078-12).

Table 1: Vendors participating in the study and type of software provided		
Vendor	**City, Country**	**Software and version**
Epsilon	Ann Arbor, Michigan, USA	EchoInsight v3.1
Esaote	Florence, Italy	XStrain2D- v5.50
Hitachi	Tokyo, Japan	2DTT Analysis v7.0a
GE Vingmed Ultrasound	Horten, Norway	EchoPac 20.1
Philips	Andover, Massachusetts, USA	QLAB aCMQ
Samsung	Seoul, South Korea	CardiacPack1.00 0615
Siemens	Mountain View, California, USA	SC2000 Workplace VVI v4.0
TOMTEC	Unterschleissheim, Germany	TOMTEC-ARENA 2DCPA
Toshiba	Otawara, Japan	ACP v. 3.2

Table 2: Normalized cross-correlation coefficient

	Noise-level							
	Low noise		Medium noise		High noise		Original image	
	mean	±SD	mean	±SD	mean	±SD	mean	±SD
Epsilon	0.67	0,16	0.71	0,15	0.71	0,18	0.63	0,26
Esaote	0.83	0,05	0.83	0,08	0.84	0,08	0.82	0,08
GE	0.78	0,05	0.78	0,07	0.67	0,21	0.79	0,03
Hitachi	0.68	0,18	0.64	0,19	0.64	0,13	0.64	0,24
Philips	0.68	0,18	0.66	0,21	0.57	0,22	0.77	0,08
Samsung	0.61	0,24	0.63	0,24	0.55	0,30	0.58	0,23
Siemens	0.83	0,07	0.80	0,09	0.79	0,05	0.84	0,08
TOMTEC	0.66	0,15	0.72	0,03	0.63	0,15	0.78	0,06
Toshiba	0.89	0,05	0.88	0,06	0.88	0,06	0.87	0,05

Normalized cross-correlation between ground truth and estimated strain values for different vendors and noise levels. The reported correlation coefficients represent average values over all models at different noise-levels. No statistically significant differences were found between the different noise levels for any of the vendors.

1

2

3

Table 3: Comparison between vendors									
	p-value for normalized cross-correlation coefficient								
	Epsilon	Esaote	GE	Hitachi	Philips	Samsung	Siemens	TOMTEC	Toshiba
Epsilon	-	0.066	0.835	0.999	1.000	0.631	0.138	1.000	0.002
Esaote	0.004	-	0.850	0.010	0.036	<0.0001	1.000	0.140	0.982
GE	<0.0001	0.001	-	0.426	0.714	0.023	0.952	0.946	0.219
Hitachi	1.000	0.005	<0.0001	-	1.000	0.947	0.024	0.990	0.000
Philips	0.091	1.000	0.005	0.179	-	0.766	0.082	1.000	0.001
Samsung	0.188	<0.0001	<0.0001	0.010	<0.0001	-	<0.0001	0.427	<0.0001
Siemens	1.000	1.000	<0.0001	1.000	1.000	0.038	-	0.263	0.920
TOMTEC	0.311	1.000	<0.0001	1.000	1.000	0.010	1.000	-	0.007
Toshiba	<0.0001	<0.0001	0.179	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-
	p-value for AUC								

4

1

Table 4: Detection of segmental positive strain (including all noise groups) for computed end-systolic strain

	AUC	*Strain cut-off value (%)*	CI lower bound	CI upper bound
Epsilon	0.80	-3.3	0.77	0.83
Esaote	0.87	-4.3	0.84	0.89
GE	0.95	-4.3	0.92	0.97
Hitachi	0.80	-3.5	0.77	0.84
Philips	0.86	-2.8	0.82	0.90
Samsung	0.72	-13.1	0.66	0.78
Siemens	0.83	-2.4	0.80	0.86
TOMTEC	0.84	-2.5	0.81	0.87
Toshiba	0.97	-2.4	0.96	0.98

Cut-off value shows the strain-value with highest sensitivity and specificity for differentiating true-positive and true-negative strain values.

1

2

Table 5: Intra-investigator variability for segmental analyses

		End-Systolic Strain Computed			
	n	Mean of Differences (%)	Std. Deviation (±%)	CI (%) Lower Bound	CI (%) Upper Bound
Epsilon	270	-0.14	1.6	-3.3	3.0
Esaote	270	-0.20	2.6	-5.2	4.8
GE	270	-0.04	1.5	-3.0	2.9
Hitachi	270	-0.06	2.5	-4.9	4.8
Philips	270	-0.17	1.8	-3.6	3.3
Samsung	270	0.24	2.4	-4.5	5.0
Siemens	270	0.14	1.6	-3.1	3.4
TOMTEC	270	-0.05	2.4	-4.8	4.7
Toshiba	270	0.14	1.6	-3.1	3.4
Mean Difference and Limits of Agreement (Bland Altman analysis) Repeated measurements on original noise level					

3

4

5

6

7

8

1

2

Table 6: Inter-investigator variability for segmental analyses					
		End-Systolic Strain Computed			
	n	Mean of Differences (%)	Std. Deviation (±%)	CI (%) Lower Bound	CI (%) Upper Bound
Epsilon	36	0,9	3,6	-6,1	8,0
Esaote	36	-3,2	5,8	-14,5	8,2
GE	36	1,2	2,4	-3,4	5,9
Hitachi	36	-0,2	4,6	-9,3	8,9
Philips	36	-0,3	5,0	-10,1	9,4
Samsung	36	-0,4	3,4	-7,1	6,3
Siemens	36	-0,3	3,4	-7,0	6,5
TOMTEC	36	0,5	3,2	-5,7	6,8
Toshiba	36	-0,1	1,8	-3,6	3,5
Mean Difference and Limits of Agreement (Bland Altman analysis) Repeated measurements on original noise level					

3

4

References:

Alessandrini M, Chakraborty B, Heyde B, Bernard O, De Craene M, Sermesant M, D'Hooge J. 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;65:411-22.

Amundsen BH, Helle-Valle T, Edvardsen T, Torp H, Crosby J, Lyseggen E, Stoylen A, Ihlen H, Lima JA, Smiseth OA, Slordahl SA. Noninvasive myocardial strain measurement by speckle tracking echocardiography: validation against sonomicrometry and tagged magnetic resonance imaging. J Am Coll Cardiol 2006;47:789-93.

Cikes M, Sutherland GR, Anderson LJ, Bijnens BH. The role of echocardiographic deformation imaging in hypertrophic myopathies. Nat Rev Cardiol 2010;7:384-96.

D'Hooge J, Barbosa D, Gao H, Claus P, Prater D, Hamilton J, Lysyansky P, Abe Y, Ito Y, Houle H, Pedri S, Baumann R, Thomas J, Badano LP, Imaging EAAITFtSD. Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data. Eur Heart J Cardiovasc Imaging 2016;17:693-701.

Dalen H, Thorstensen A, Aase SA, Ingul CB, Torp H, Vatten LJ, Stoylen A. Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: the HUNT study in Norway. Eur J Echocardiogr 2010;11:176-83.

Duchateau N, Loncaric F, Cikes M, Doltra A, Sitges M, Bijnens B. Variability in the Assessment of Myocardial Strain Patterns: Implications for Adequate Interpretation. Ultrasound Med Biol 2020;46:244-54.

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. Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography 2015;28:1171-81, e2.

Friedberg MK, Mertens L. Tissue velocities, strain, and strain rate for echocardiographic assessment of ventricular function in congenital heart disease. Eur J Echocardiogr 2009;10:585-93.

Grenne B, Eek C, Sjoli B, Dahlslett T, Uchto M, Hol PK, Skulstad H, Smiseth OA, Edvardsen T, Brunvand H. Acute coronary occlusion in non-ST-elevation acute coronary syndrome: outcome and early identification by strain echocardiography. Heart 2010;96:1550-6.

Grue JF, Storve S, Dalen H, Salvesen O, Mjolstad OC, Samstad SO, Torp H, Haugen BO. Automatic Measurements of Mitral Annular Plane Systolic Excursion and Velocities to Detect Left Ventricular Dysfunction. Ultrasound Med Biol 2018;44:168-76.

Hanekom L, Jenkins C, Jeffries L, Case C, Mundy J, Hawley C, Marwick TH. Incremental value of strain rate analysis as an adjunct to wall-motion scoring for assessment of myocardial viability by dobutamine echocardiography: a follow-up study after revascularization. Circulation 2005;112:3892-900.

Haugaa KH, Goebel B, Dahlslett T, Meyer K, Jung C, Lauten A, Figulla HR, Poerner TC, Edvardsen T. Risk assessment of ventricular arrhythmias in patients with nonischemic dilated cardiomyopathy by strain echocardiography. J Am Soc Echocardiogr 2012;25:667-73.

Hsiao JF, Koshino Y, Bonnichsen CR, Yu Y, Miller FA, Jr., Pellikka PA, Cooper LT, Jr., Villarraga HR. Speckle tracking echocardiography in acute myocarditis. Int J Cardiovasc Imaging 2013;29:275-84.

Leitman M, Lysyansky P, Sidenko S, Shir V, Peleg E, Binenbaum M, Kaluski E, Krakover R, Vered Z. Two-dimensional strain-a novel software for real-time quantitative echocardiographic assessment of myocardial function. J Am Soc Echocardiogr 2004;17:1021-9.

Marchesseau S, Delingette H, Sermesant M, Sorine M, Rhode K, Duckett SG, Rinaldi CA, Razavi R, Ayache N. Preliminary specificity study of the Bestel-Clement-Sorine electromechanical

model of the heart using parameter calibration from medical images. J Mech Behav Biomed Mater 2013;20:259-71.
Marciniak A, Eroglu E, Marciniak M, Sirbu C, Herbots L, Droogne W, Claus P, D'Hooge J, Bijnens B, Vanhaecke J, Sutherland GR. The potential clinical role of ultrasonic strain and strain rate imaging in diagnosing acute rejection after heart transplantation. Eur J Echocardiogr 2007;8:213-21.
Menet A, Bernard A, Tribouilloy C, Leclercq C, Gevaert C, Guyomar Y, Guerbaai RA, Delelis F, Castel AL, Graux P, Ennezat PV, Donal E, Marechaux S. Clinical significance of septal deformation patterns in heart failure patients receiving cardiac resynchronization therapy. Eur Heart J Cardiovasc Imaging 2017;18:1388-97.
Mirea O, Pagourelias ED, Duchenne J, Bogaert J, Thomas JD, Badano LP, Voigt JU, Force EA-A-IST. Intervendor Differences in the Accuracy of Detecting Regional Functional Abnormalities: A Report From the EACVI-ASE Strain Standardization Task Force. JACC Cardiovasc Imaging 2018;11:25-34.
Mirea O, Pagourelias ED, Duchenne J, Bogaert J, Thomas JD, Badano LP, Voigt JU, Force EA-A-IST. Variability and Reproducibility of Segmental Longitudinal Strain Measurement: A Report From the EACVI-ASE Strain Standardization Task Force. JACC Cardiovasc Imaging 2018;11:15-24.
Miyasaka Y, Haiden M, Kamihata H, Nishiue T, Iwasaka T. Usefulness of strain rate imaging in detecting ischemic myocardium during dobutamine stress. Int J Cardiol 2005;102:225-31.
Parsai C, Bijnens B, Sutherland GR, Baltabaeva A, Claus P, Marciniak M, Paul V, Scheffer M, Donal E, Derumeaux G, Anderson L. Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms. Eur Heart J 2009;30:940-9.
Phelan D, Collier P, Thavendiranathan P, Popovic ZB, Hanna M, Plana JC, Marwick TH, Thomas JD. Relative apical sparing of longitudinal strain using two-dimensional speckle-tracking echocardiography is both sensitive and specific for the diagnosis of cardiac amyloidosis. Heart 2012;98:1442-8.
Rosner A, Avenarius D, Malm S, Iqbal A, Bijnens B, Schirmer H. Severe regional myocardial dysfunction by stress echocardiography does not predict the presence of transmural scarring in chronic coronary artery disease. Eur Heart J Cardiovasc Imaging 2015;16:1074-81.
Rosner A, Barbosa D, Aarsaether E, Kjonas D, Schirmer H, D'Hooge J. The influence of frame rate on two-dimensional speckle-tracking strain measurements: a study on silico-simulated models and images recorded in patients. Eur Heart J Cardiovasc Imaging 2015;16:1137-47.
Stoylen A, Heimdal A, Bjornstad K, Wiseth R, Vik-Mo H, Torp H, Angelsen B, Skjaerpe T. Strain rate imaging by ultrasonography in the diagnosis of coronary artery disease. J Am Soc Echocardiogr 2000;13:1053-64.
Tabassian M, Alessandrini M, Herbots L, Mirea O, Pagourelias ED, Jasaityte R, Engvall J, De Marchi L, Masetti G, D'Hooge J. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. Int J Cardiovasc Imaging 2017;33:1159-67.
Thomas JD, Badano LP. EACVI-ASE-industry initiative to standardize deformation imaging: a brief update from the co-chairs. Eur Heart J Cardiovasc Imaging 2013;14:1039-40.
Thorstensen A, Amundsen BH, Dalen H, Hala P, Kiss G, Aase SA, Torp H, Stoylen A. Strain rate imaging combined with wall motion analysis gives incremental value in direct quantification of myocardial infarct size. Eur Heart J Cardiovasc Imaging 2012;13:914-21.
Unlu S, Duchenne J, Mirea O, Pagourelias ED, Bezy S, Cvijic M, Beela AS, Thomas JD, Badano LP, Voigt JU, Force E-AIST. Impact of apical foreshortening on deformation measurements: a report from the EACVI-ASE Strain Standardization Task Force. Eur Heart J Cardiovasc Imaging 2020;21:337-43.
Voigt JU, Pedrizzetti G, Lysyansky P, Marwick TH, Houle H, Baumann R, Pedri S, Ito Y, Abe Y, Metz S, Song JH, Hamilton J, Sengupta PP, Kolias TJ, d'Hooge J, Aurigemma GP, Thomas JD, Badano

LP. 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.

Weidemann F, Rummey C, Bijnens B, Stork S, Jasaityte R, Dhooge J, Baltabaeva A, Sutherland G, Schulz JB, Meier T, Mitochondrial Protection with Idebenone in Cardiac or Neurological Outcome study g. The heart in Friedreich ataxia: definition of cardiomyopathy, disease severity, and correlation with neurological symptoms. Circulation 2012;125:1626-34.

Zhang Y, Chan AK, Yu CM, Yip GW, Fung JW, Lam WW, So NM, Wang M, Wu EB, Wong JT, Sanderson JE. Strain rate imaging differentiates transmural from non-transmural myocardial infarction: a validation study using delayed-enhancement magnetic resonance imaging. J Am Coll Cardiol 2005;46:864-71.

Zwanenburg JJ, Gotte MJ, Kuijer JP, Hofman MB, Knaapen P, Heethaar RM, van Rossum AC, Marcus JT. Regional timing of myocardial shortening is related to prestretch from atrial contraction: assessment by high temporal resolution MRI tagging in humans. Am J Physiol Heart Circ Physiol 2005;288:H787-94.

Aarsaether E, Rosner A, Straumbotn E, Busund R. Peak longitudinal strain most accurately reflects myocardial segmental viability following acute myocardial infarction - an experimental study in open-chest pigs. Cardiovasc Ultrasound 2012;10:23.

Figure legends

Figure 1: Image of different models in the three standard projections. Blue indicates infarcted areas with end-systolic stretching.

Figure 2: End-systolic (ES) computed strain-values of the models for the different vendors divided into stretching (ES strain >0%) and shortening segments (ES strain <0%) based on the ground truth value.

Figure 3: Segmental absolute errors for different noise levels and vendors. Black bars indicate confidence intervals. *significant differences between original and at least one noise level.

Figure 4: Bland-Altman plots for computed values (COMP) vs. ground truth (GT) for the different vendors. SD: standard deviation.

Figure 5: ROC curves for segmental analysis and the differentiation between segmental shortening and stretching. Analysis for the different vendors.

Supplement Figure 1: Regression plots for each vendor between ground truth (GT) and computed end-systolic (ES) strain in the original image without noise.

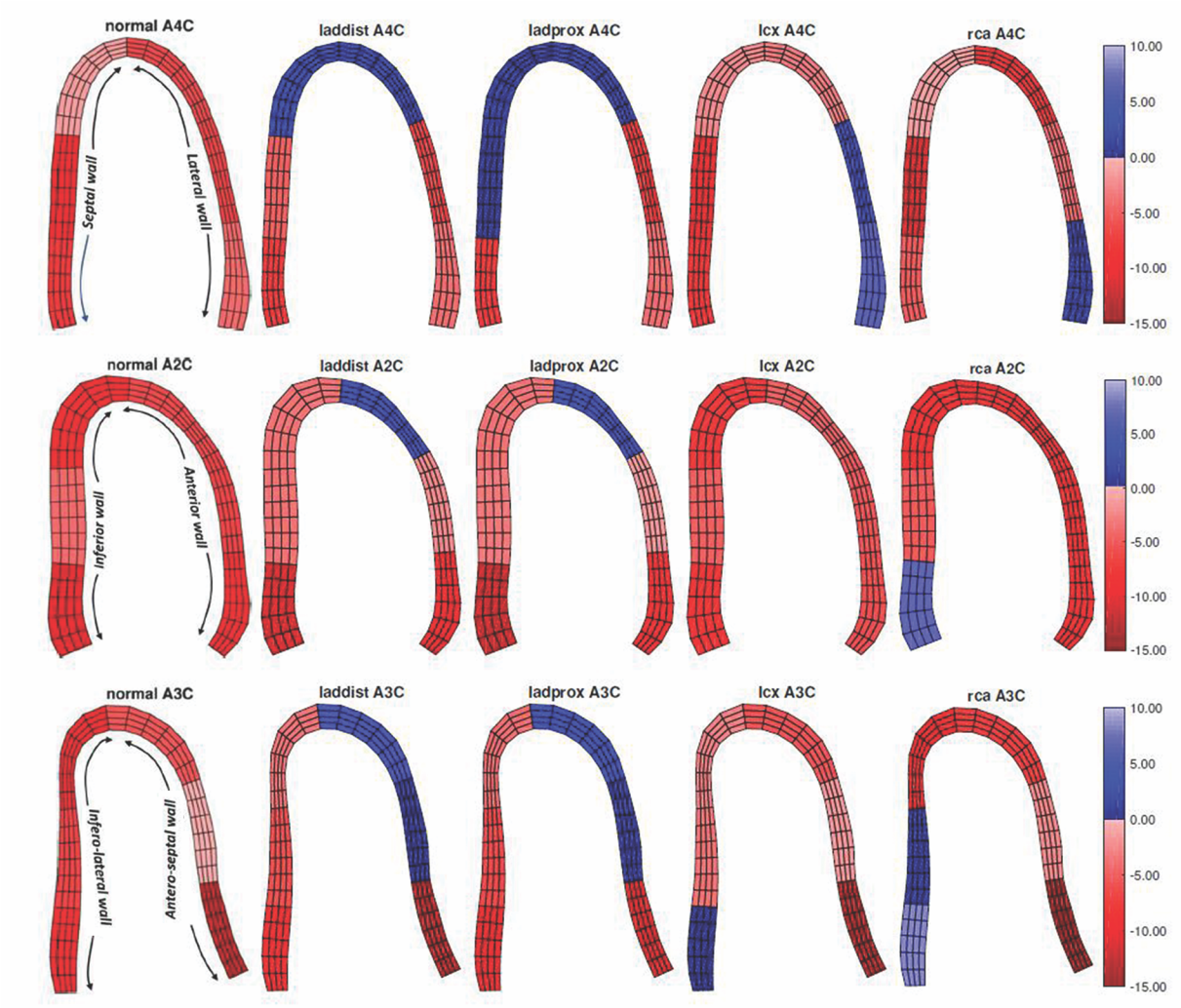

normal A4C
laddist A4C
ladprox A4C
lcx A4C
rca A4C
Septal wall
Lateral wall
normal A2C
laddist A2C
ladprox A2C
lcx A2C
rca A2C
Inferior wall
Anterior wall
normal A3C
laddist A3C
ladprox A3C
lcx A3C
rca A3C
Infero-lateral wall
Antero-septal wall
10.00
5.00
0.00
-5.00
-10.00
-15.00

Figure 1

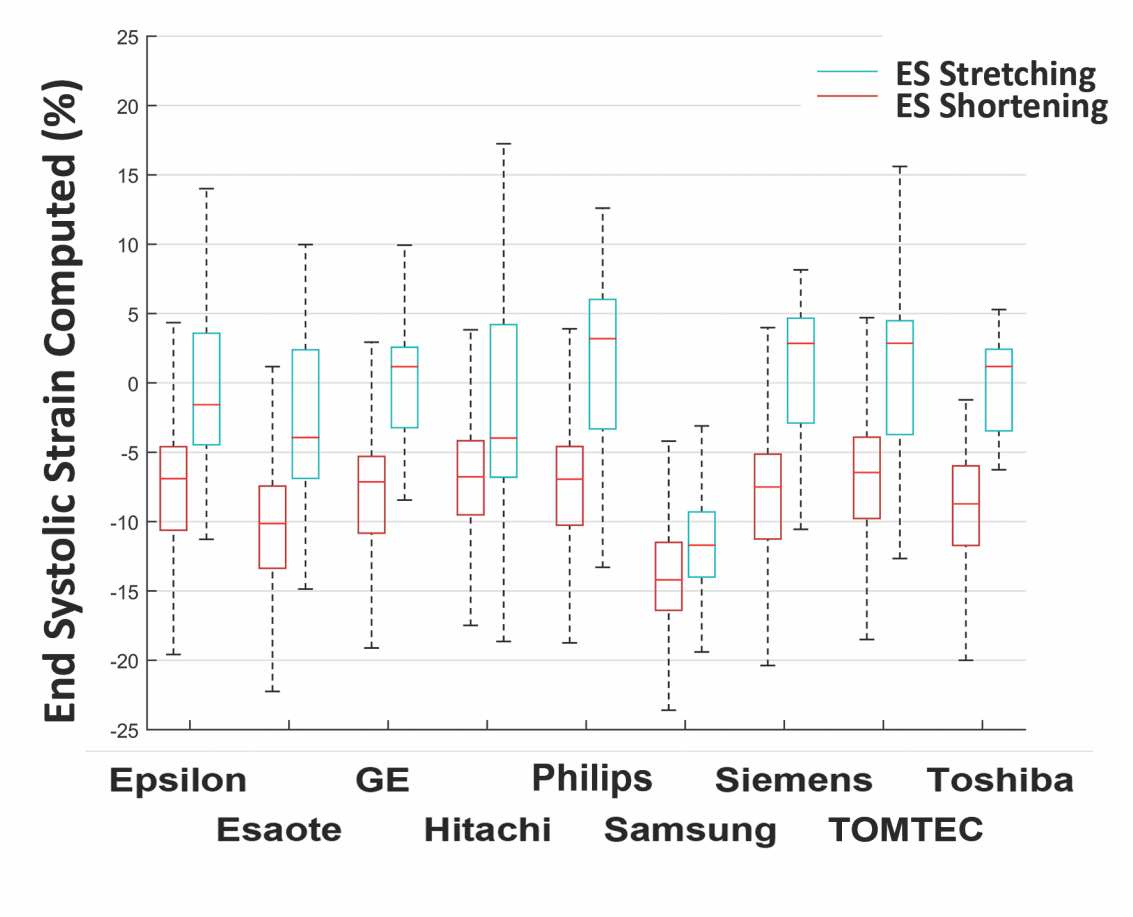

End Systolic Strain Computed (%)
25
20
15
10
5
0
-5
-10
-15
-20
-25
ES Stretching
ES Shortening
Epsilon
GE
Philips
Siemens
Toshiba
Esaote
Hitachi
Samsung
TOMTEC

Figure 2

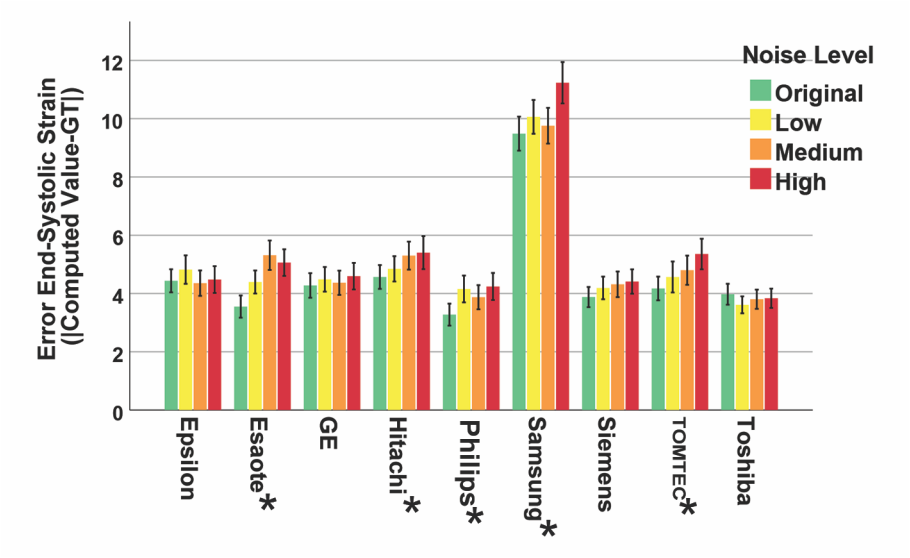
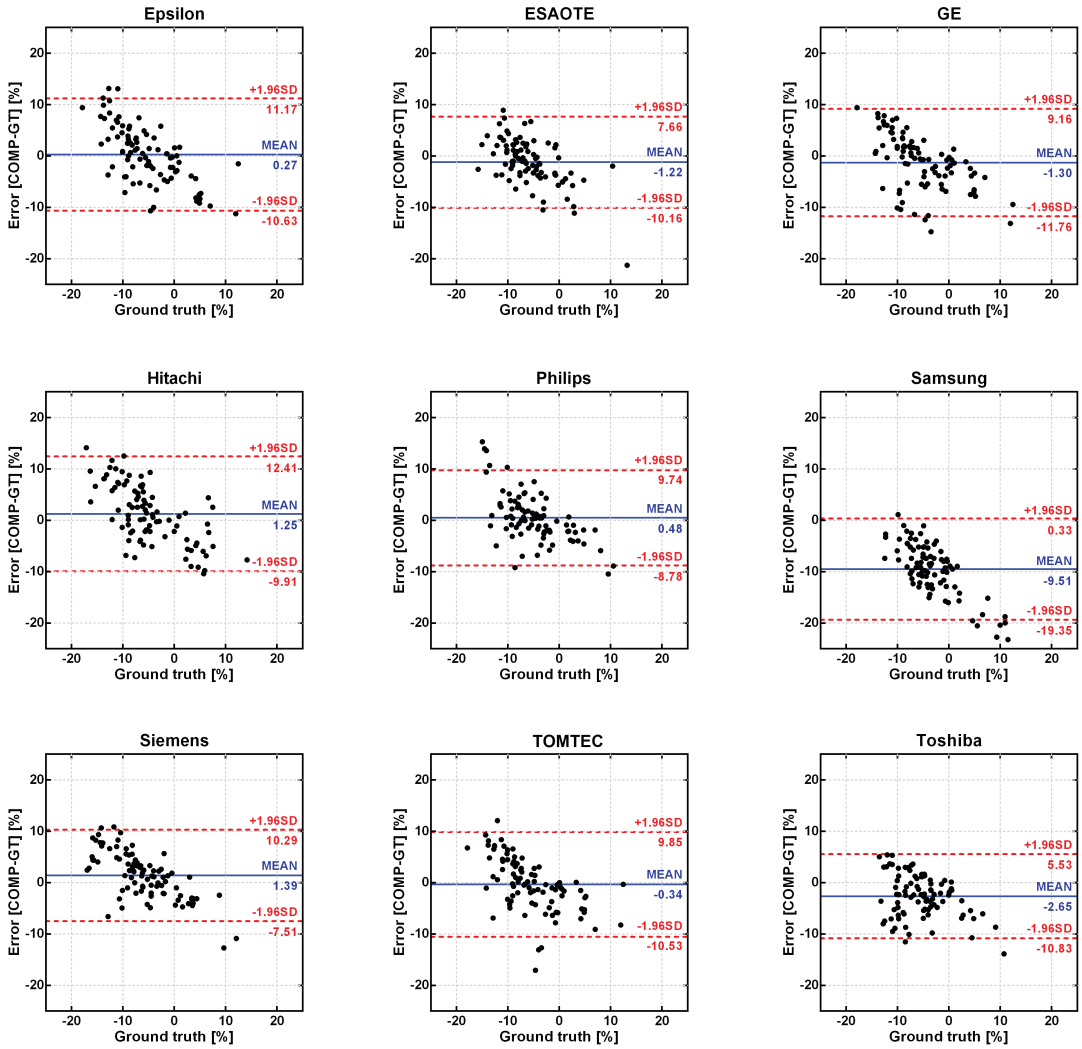

Error End-Systolic Strain
(|Computed Value-GT|)
12
10
8
6
4
2
0
Noise Level
Original
Low
Medium
High
Epsilon
Esaote*
GE
Hitachi*
Philips*
Samsung*
Siemens
TOMTEC*
Toshiba

Figure 3


Epsilon
+1.96SD
11.17
MEAN
0.27
-1.96SD
-10.63
ESAOTE
+1.96SD
7.66
MEAN
-1.22
-1.96SD
-10.16
GE
+1.96SD
9.16
MEAN
-1.30
-1.96SD
-11.76
Hitachi
+1.96SD
12.41
MEAN
1.25
-1.96SD
-9.91
Philips
+1.96SD
9.74
MEAN
0.48
-1.96SD
-8.78
Samsung
+1.96SD
0.33
MEAN
-9.51
-1.96SD
-19.35
Siemens
+1.96SD
10.29
MEAN
1.39
-1.96SD
-7.51
TOMTEC
+1.96SD
9.85
MEAN
-0.34
-1.96SD
-10.53
Toshiba
+1.96SD
5.53
MEAN
-2.65
-1.96SD
-10.83
Error [COMP-GT] [%]
Ground truth [%]

1

2 Figure 4

3

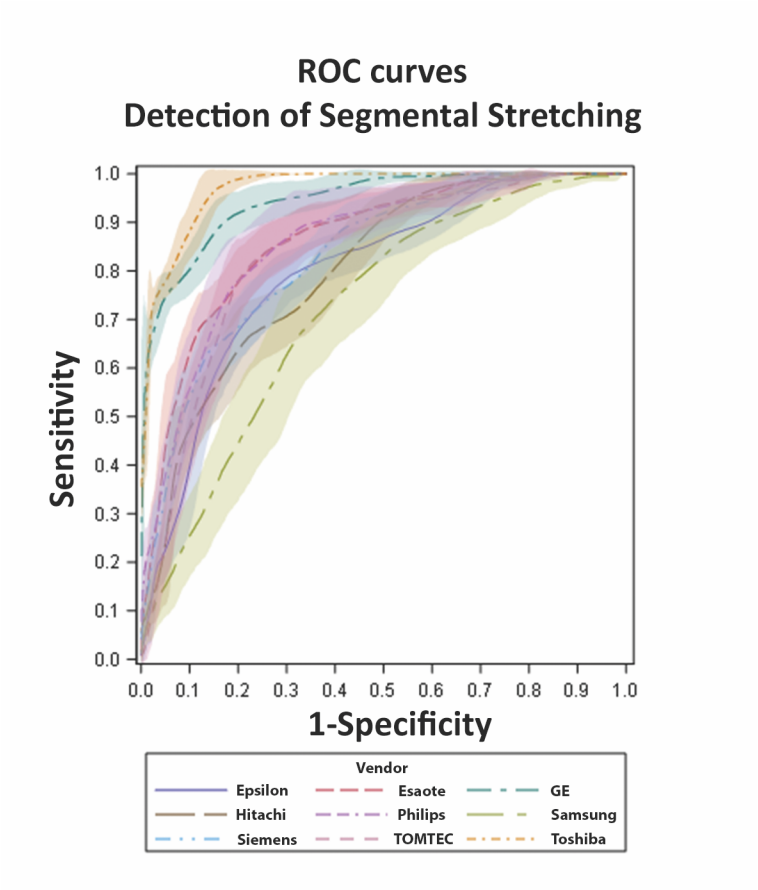

ROC curves
Detection of Segmental Stretching
Sensitivity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
1-Specificity
Vendor
Epsilon
Esaote
GE
Hitachi
Philips
Samsung
Siemens
TOMTEC
Toshiba

Figure 5