Artificial intelligence based left ventricular ejection fraction and global longitudinal strain in cardiac amyloidosis

Cotella, J., Slivnick, Jeremy A., Sanderson, Emily, Singulane, Cristiane, O'Driscoll, Jamie, Asch, Federico M., Addetia, K., Woodward, Gary and Lang, Roberto M.

Artificial Intelligence Based Left Ventricular Ejection Fraction and Global Longitudinal Strain in Cardiac Amyloidosis

Juan I. Cotella MD[1], Jeremy A. Slivnick MD[1], Emily Sanderson MSc[2], Cristiane Singulane MD[1], Jamie O'Driscoll PhD [2,3,4], Federico M. Asch MD[5], Karima Addetia MD[1], Gary Woodward PhD[2], Roberto M. Lang MD[1]

[1] University of Chicago Medical Center, Chicago, Illinois, United States;

[2] Ultromics Ltd, 4630 Kingsgate, Cascade Way, Oxford Business Park South, Oxford, OX4 2SU;

[3] Canterbury Christ Church University, School of Psychology and Life Sciences, North Holmes Road, Kent, CT1 1QT;

[4] Department of Cardiology, St George's University Hospitals NHS Foundation Trust, Blackshaw Road, Tooting, London, SW17 0QT;

[5] Cardiovascular Core Labs, MedStar Health Research Institute, Washington DC, United States.

Running Title: **AI-derived LVEF and GLS in Cardiac Amyloidosis**

-The data that support the findings of this study are available from the corresponding author upon reasonable request

-No funding was received for this manuscript

Correspondence to:
Roberto M Lang, MD
The University of Chicago Medicine
5758 S. Maryland Avenue, MC 9067, DCAM 5509
Chicago, IL 60637

Abstract

Background: Assessment of left ventricular ejection fraction (LVEF) and global longitudinal strain (GLS) plays a key role in the diagnosis of cardiac amyloidosis (CA). However, manual measurements are time consuming and prone to variability. We aimed to assess whether fully automated artificial intelligence (AI) calculation of LVEF and GLS provide similar estimates and can identify abnormalities in agreement with conventional manual methods, in patients with pre-clinical and clinical CA.

Methods: We identified 51 patients (age 80±10 years, 53% male) with confirmed CA according to guidelines, who underwent echocardiography before and/or at the time of CA diagnosis (median (IQR) time between observations 3.87 (1.93, 5.44) yrs). LVEF and GLS were quantified from the apical 2- and 4-chamber views using both manual and fully automated methods (EchoGo Core 2.0, Ultromics). Inter-technique agreement was assessed using linear regression and Bland-Altman analyses and two-way ANOVA. The diagnostic accuracy and time for detecting abnormalities (defined as LVEF ≤50% and GLS≥-15.1%, respectively) using AI was assessed by comparisons to manual measurements as a reference.

Results: There were no significant differences in manual and automated LVEF and GLS values in either pre-CA (p=0.791 and p=0.105, respectively) or at diagnosis (p=0.463 and p=0.722). The two methods showed strong correlation on both the pre-CA (r=0.78 and r=0.83) and CA echoes (r=0.74 and r=0.80) for LVEF and GLS, respectively. The sensitivity and specificity of AI-derived indices for detecting abnormal LVEF were 83% and 86%, respectively, in the pre-CA echo and 70% and 79% at CA diagnosis. The sensitivity and specificity of AI-derived indices for detecting abnormal GLS was 82% and 86% in the pre-CA echo and 100% and 67% at the time of CA diagnosis. There was no significant difference in the relationship between LVEF (p=0.99) and GLS (p=0.19) and time to abnormality between the two methods.

Conclusion: Fully automated AI-calculated LVEF and GLS are comparable to manual measurements in patients pre-CA and at the time of CA diagnosis. The widespread implementation of automated LVEF and GLS may allow for more rapid assessment in different disease states with comparable accuracy and reproducibility to manual methods.

Key words: Amyloidosis, Artificial intelligence, ejection fraction, global longitudinal strain, deep learning.

Introduction

Cardiac amyloidosis (CA) is a rapidly progressive infiltrative cardiomyopathy, which carries a poor prognosis despite disease-modifying therapies (1). The major amyloidosis subtypes are acquired monoclonal immunoglobulin light chain amyloidosis (AL) and transthyretin (ATTR) amyloidosis, including mutant (ATTRm) and wild-type (ATTRwt), which are caused by misfolding of the transthyretin protein. Myocardial involvement occurs from extracellular deposition and accumulation of β-pleated insoluble fibrillar proteins in the heart, leading to ventricular hypertrophy, stiffening and impaired relaxation. As such, CA is often categorised as a restrictive cardiomyopathy (2) and commonly progresses to heart failure from mechanical, biochemical and electrical dysfunction (3).

CA is clinically challenging and early detection is extremely important (2). Echocardiography is normally the first line imaging tool for the investigation of CA (4), and knowledge of echocardiographic features of CA is essential for accurate identification and differential diagnosis (5). Previous research has extensively demonstrated the clinical utility of 2D and Doppler echocardiography, including global longitudinal strain (GLS) as predictors of poor outcome in patients with CA (4,6-13). However, at present, quantifiable parameters such as left ventricular (LV) ejection fraction (EF), volume and GLS require manual contouring of the LV endocardium, and are subject to significant inter- and intra-operator variability (14), which can affect the accuracy of making a diagnosis or decisions about treatment.

Artificial intelligence (AI) algorithms can reduce observer variability (15). Recently, an AI image-processing pipeline that can automatically analyze echocardiograms was utilized in patients with COVID-19 infection (16,17). The platform calculates LV size and systolic function, including LVEF and GLS, without requirement for any input from the clinician. In the current study, we aimed to assess whether this approach for a fully automated, AI-based

calculation of LVEF and GLS can provide similar estimates to conventional methodology and identify abnormalities in agreement with manual methods, in patients with pre-clinical and clinical CA.

Methods

AI-algorithm training and validation

An application developed by Ultromics (EchoGo Core 2.0) was used to provide automatic quantification of commonly measured echocardiographic metrics of LV function (Figure 1). First, to identify a heart cycle, the operator selected a contoured image clip for each view at end-diastolic (ED) and end-systolic (ES) frames. Then, the auto-viewer classifier (AVC) tool automatically identified the required views from an echocardiogram input. The AVC convolutional neural network (CNN) consists of 10 two-dimensional (2D) convolutional layers for extracting image features. Each convolution uses a 3x3 kernel and it is followed by batch normalization and a rectified linear unit activation function. After every two convolutions, the feature map is down-sampled by a factor of 2 using a max pooling layer. Dropout was used for reducing overfitting and applied to the last convolutional layer as well as the last fully connected layer. The output of the fully connected layer was fed into a Softmax layer to represent probabilities of the different echocardiographic views. The prediction probability of a video was calculated by taking the mean of probabilities from all its frames. Each echocardiogram video was predicted as one of the classification classes based on the highest Softmax probability.

The AVC was then trained to identify apical-2-chamber (A2C), apical-3-chamber (A3C) and apical-4-chamber (A4C) views using 27334, 27932, 27004 frames, respectively, and then validated on an independent dataset of 8178, 8662 and 8397 frames, respectively. Validation accuracy was 0.95, 0.96 and 1.00 for the A2C, A3C and A4C views, respectively.

The LV segmentation method employed in all auto-contouring models is based on contour points regression. The LV segmentation method directly outputs the likelihood of 2D coordinates of the contour landmark points. Images are passed through a 2D U-Net model, that learns abstract representations of the given input. The output of the U-Net is treated as histograms of the location-likelihood of the segmentation landmark points. These histograms are subsequently passed through a differentiable spatial to numerical transform (DSNT) that converts the histograms into 2D coordinates that correspond to the contour points. The auto-contouring framework was developed using a CNN, which was trained using labelled data of contours of the endocardial border, manually drawn by experienced echocardiographers. The AVC was trained to contour A2C, A3C and A4C views using 3064, 4822, 4059 ED or ES frames (respectively) and validated on an independent dataset of 728, 862 and 985 ED or ES frames (respectively). Validation coefficients were 0.892, 0833 and 0.912 for the A2C, A3C and A4C views, respectively.

Study Population

We retrospectively studied 51 patients (age 80±10 years, 53% male) with confirmed CA (AL 21.6%, ATTR 70.6%, undefined 7.8%) according to current guidelines (18), who underwent transthoracic echocardiography (TTE) pre-CA diagnosis and/or at the time of CA diagnosis and had complete data for both AI and manual measurements for at least one timepoint for LVEF and/or GLS (Table 1). Once the CA diagnosis had been made, we identified patients who had undergone a TTE study prior to this diagnosis, and included these patients in our analysis, as the “pre-CA diagnosis” echo (median [IQR] time between studies 3.87 [1.93, 5.44] years). Basic clinical information and DICOM cardiac ultrasound images were collected from the PACS systems, deidentified, and transferred via a web-based system (Ultromics, Oxford, UK) for central analysis by a core group lead by the principal investigators. Image transfer was facilitated by a two-step anonymization process to a cloud-

based image analysis software. The study was approved by the University of Chicago institutional review board.

Image Analysis

Analysis was performed through commercially available AI algorithms based on machine learning (Ultromics, EchoGo Core 2.0), which generated an initial LV EF, end-systolic and end-diastolic volumes (LVESV and LVEDV, respectively), and GLS. Automated measures were compared against conventional manual measurements, which were performed by certified echocardiographers from the principal investigator's core laboratory, blinded to any clinical data and the results of the automated analysis. All LV volume and EF measurements were performed using endocardial tracings and the biplane method of disks (modified Simpson's rule) (19). Only cases with acceptable-quality LV views were included, defined as lack of apical foreshortening with adequate visualization of all segments in the apical 4-chamber (4CH) view. From the initial cohort of 59 studies, 8 were excluded due to inadequate image quality. Longitudinal strain was calculated as the average of all available segments from the 4CH and 2-chamber (2CH) views, since a long-axis view was not obtained in the vast majority of cases (Figure 2). Cutoffs for mildly, moderately, and severely reduced LVEF were determined by the 2015 American Society of Echocardiography (ASE)/European Association of Cardiovascular Imaging (EACVI) guidelines for cardiac chamber quantification (19). Normal GLS was defined as <–19%, mildly abnormal GLS was defined as >–19% and <–15%, and severely abnormal GLS was defined as >–15% (20).

Statistical Analysis

Paired t-tests were performed to compare differences in measurements between manual and AI methods at the pre-CA timepoint and at CA diagnosis. Bland-Altman analysis was conducted to compare methods at each time point. Pearson's correlation coefficients were

used to establish the strength of relationship between measurements by the two methods. Two-way ANOVA was used to compare the change in measurement from pre-CA to diagnosis across methods.

The measurements of LV EF and GLS were dichotomized using cut-offs of EF<50% and GLS>-15.1%. Sensitivity and specificity of AI measurements were calculated at each matched timepoint using the manual measurements as the reference standard.
The longitudinal dichotomized data were used to determine whether there was an association between the method of measurement and the "time to event", which was defined as the time between seeing a value indicative of CA in the early timepoint and the official diagnosis. Kaplan-Meier curves and log-rank tests were used to compare these times to event.

Results

There was no significant difference in LV EF measurements pre-CA (57.0±12.6 vs 57.3±10.8, *p*=0.791) and at the time of CA diagnosis (46.5±16.1 vs 45.2±12.6, *p*=0.463) between manual and AI derived values, respectively. There was also no significant difference between the change in LV EF from pre-CA to CA diagnosis (8.0±20.4 vs. 9.4±18.8; *p*=0.603) for manual and AI derived measures, respectively. Bland-Altman analysis showed that 91.4% of pairs fall within the limits of agreement pre-CA and 100% of pairs fall within the limits of agreement at time of diagnosis (Figure 3A-B). Manual and AI derived measures of LVEF demonstrated good correlations for both the pre-CA (r=0.78) and CA echo (r=0.74) (Figure 4). Using the manual measurements as the reference standard, the sensitivity and specificity of AI-derived indices for detecting abnormal LVEF were 83%, and 86% in the pre-CA echo, and 70% and 79% at the time of CA diagnosis, respectively. When looking at the time

between pre-CA measurements and diagnosis, there was no significant difference in the time to abnormality between the two methods (p=0.990).

There was no significant difference in GLS values of pre-CA (-17.7±5.0 vs -16.9±4.6, *p*=0.105) and at the time of CA diagnosis (-12.7±4.5 vs -12.9±4.7, *p*=0.722) between manual and AI derived measurements, respectively. Bland-Altman analysis shows that 91.9% of pairs fall within the limits of agreement pre-CA and 96.1% of pairs fall within the limits of agreement at the time of diagnosis (Figure 5 A-B). There was also no significant difference between the change in GLS from pre-CA to CA diagnosis (-2.9±6.3 vs. -2.3±7.0; *p*=0.548) for manual and AI derived measures, respectively. Manual and AI derived measures of GLS demonstrated strong correlations for both the pre-CA (r=0.83) and CA echo (r=0.80) (Figure 6). Using the manual measurements as the reference standard, the sensitivity and specificity of AI-derived indices for detecting abnormal GLS were 82%, and 86% in the pre-CA echo, and 100% and 67% at the time of CA diagnosis, respectively.

When looking at the time between pre-CA measurements and diagnosis, there was no significant difference in the time to abnormality between the two methods (p=0.192).

Discussion

This study demonstrated that quantitative, automated measurements of LVEF and GLS are highly accurate, and allow sensitive and specific detection of abnormalities, when compared to the conventional manual analysis, in patients undergoing echocardiographic assessment pre-CA diagnosis and at the time of CA diagnosis. The hypothesis and rationale for choosing 2 separate time points, was that use of the AI algorithm in these patients could allow the detection of CA related abnormalities even before the diagnosis was made according to current recommendations. This novel approach could have the potential of providing

clinicians with automated quantification of LV function and mechanics, that could help raise suspicion for CA in a pre-clinical scenario, ultimately leading to the diagnosis at earlier stages of the disease. Our results demonstrate that AI-calculated LVEF and GLS can be highly complementary to clinical practice, by producing representative data across longitudinal observations without additional time, resources, or expertise from the clinician. Previous research has detailed that conventional echocardiographic parameters have low accuracy in the diagnosis of CA, mostly driven by poor sensitivity (12). Nevertheless, GLS and patterns of myocardial deformation demonstrated higher sensitivity and specificity for the detection of CA. Indeed, Pagourelias et al. (12) demonstrated that the LVEF/GLS ratio was the best discriminator for CA, even in challenging subgroups. However, our LVEF/GLS ratio demonstrated poor sensitivity (25%) with good specificity (90%), suggesting that it could possibly be used to rule in potential amyloidosis cases.

Recently, Cohen et al. (4) demonstrated the importance of GLS in the assessment of CA. Baseline GLS was an independent predictor of survival (above standard biomarkers) and a change in GLS of -2.0% at 12 and 24-months following clinical intervention improved long-term survival. These results have significant clinical implications, particularly surrounding the importance of accurate and reliable measures of cardiac imaging parameters, since CA patients usually undergo serial echocardiographic assessments. In current practice, LVEF and GLS is manually contoured and wide variation in inter- and intra-observer variability has been reported. Indeed, an inter-observer variability of 5.4% to 8.6% and intra-observer variability of 4.9% to 7.3% has been reported for GLS, which is lower than that reported for LVEF (21). This level of variation in manual contouring, which in clinical practice, often exceeds the reported values, may have implications for patients, especially surrounding the clinically relevant change of -2.0% in GLS (4).

Interestingly, despite the strong correlation between manual and AI derived measures of GLS in both the pre-CA and CA studies, this parameter showed a lower specificity at the time of diagnosis (67%) than at the pre-CA time (86%). The reason of these findings is unclear but might be related to differences in image quality between the studies. Moreover, the small sample size, the unavailability of adequate A3C views to include in the LV GLS analysis and the retrospective nature of this study should be taken into consideration when interpreting our findings.

The AI pipeline performs automated measures of LVEF and GLS and compares to manual measures performed at a center of excellence. This workflow also offers additional advantages over current practice, such as the elimination of training to perform GLS and LVEF measures, a reduction in the time needed to perform these measures routinely, and the inter-vendor variability (21).

Conclusion

AI-calculated LVEF and GLS are accurate in both pre-clinical CA and the time of CA diagnosis. The widespread implementation of automated LVEF and GLS may allow for more rapid assessment in different disease states with comparable accuracy to manual methods.

References

1. Coelho T, Adams D, Silva A et al. Safety and efficacy of RNAi therapy for transthyretin amyloidosis. N Engl J Med 2013;369:819-29.
2. Guan J, Mishra S, Falk RH, Liao R. Current perspectives on cardiac amyloidosis. Am J Physiol Heart Circ Physiol 2012;302:H544-52.
3. Muchtar E, Blauwet LA, Gertz MA. Restrictive Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy. Circ Res 2017;121:819-837.
4. Cohen OC, Ismael A, Pawarova B et al. Longitudinal strain is an independent predictor of survival and response to therapy in patients with systemic AL amyloidosis. Eur Heart J 2022;43:333-341.
5. Phelan D, Collier P, Thavendiranathan P et al. 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.
6. Barros-Gomes S, Williams B, Nhola LF et al. Prognosis of Light Chain Amyloidosis With Preserved LVEF: Added Value of 2D Speckle-Tracking Echocardiography to the Current Prognostic Staging System. JACC Cardiovasc Imaging 2017;10:398-407.
7. Bellavia D, Pellikka PA, Al-Zahrani GB et al. Independent predictors of survival in primary systemic (Al) amyloidosis, including cardiac biomarkers and left ventricular strain imaging: an observational cohort study. J Am Soc Echocardiogr 2010;23:643-52.
8. Buss SJ, Emami M, Mereles D et al. Longitudinal left ventricular function for prediction of survival in systemic light-chain amyloidosis: incremental value compared with clinical and biochemical markers. J Am Coll Cardiol 2012;60:1067-76.
9. Cueto-Garcia L, Reeder GS, Kyle RA et al. Echocardiographic findings in systemic amyloidosis: spectrum of cardiac involvement and relation to survival. J Am Coll Cardiol 1985;6:737-43.
10. Klein AL, Hatle LK, Taliercio CP et al. Prognostic significance of Doppler measures of diastolic function in cardiac amyloidosis. A Doppler echocardiography study. Circulation 1991;83:808-16.
11. Koyama J, Falk RH. Prognostic significance of strain Doppler imaging in light-chain amyloidosis. JACC Cardiovasc Imaging 2010;3:333-42.
12. Pagourelias ED, Mirea O, Duchenne J et al. Echo Parameters for Differential Diagnosis in Cardiac Amyloidosis: A Head-to-Head Comparison of Deformation and Nondeformation Parameters. Circ Cardiovasc Imaging 2017;10:e005588.
13. Quarta CC, Solomon SD, Uraizee I et al. Left ventricular structure and function in transthyretin-related versus light-chain cardiac amyloidosis. Circulation 2014;129:1840-9.
14. Marwick TH. Ejection Fraction Pros and Cons: JACC State-of-the-Art Review. J Am Coll Cardiol 2018;72:2360-2379.
15. Tromp J, Seekings PJ, Hung CL et al. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. Lancet Digit Health 2022;4:e46-e54.
16. Karagodin I, Carvalho Singulane C, Woodward GM et al. Echocardiographic Correlates of In-Hospital Death in Patients with Acute COVID-19 Infection: The World Alliance

Societies of Echocardiography (WASE-COVID) Study. J Am Soc Echocardiogr 2021;34:819-830.

17. Karagodin I, Singulane CC, Descamps T et al. Ventricular Changes in Patients with Acute COVID-19 Infection: Follow-up of the World Alliance Societies of Echocardiography (WASE-COVID) Study. J Am Soc Echocardiogr 2021.
18. Garcia-Pavia P, Rapezzi C, Adler Y et al. Diagnosis and treatment of cardiac amyloidosis: a position statement of the ESC Working Group on Myocardial and Pericardial Diseases. Eur Heart J 2021;42:1554-1568.
19. 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. J Am Soc Echocardiogr 2015;28:1-39 e14.
20. Yang H, Wright L, Negishi T, Negishi K, Liu J, Marwick TH. Research to Practice: Assessment of Left Ventricular Global Longitudinal Strain for Surveillance of Cancer Chemotherapeutic-Related Cardiac Dysfunction. JACC Cardiovasc Imaging 2018;11:1196-1201.
21. 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, e2.

Table 1. Basic demographic data of 51 study patients who had complete data for both AI and manual measurements for at least one time point for LVEF and/or GLS (n = 51 patients)

	N (%)
Gender (Male)	27 (51)
Type of CA	**AL-CA** 11 (21.6) **ATTR + ATTR-wt** 36 (70.6) **Undefined** 4 (7.8)
Race	**Black/African American** 45 (88.2) **White** 7 (11.8)
	Mean ± SD
Age (years)	80.2 ± 9.7
Height (cm)	171 ± 10.3
Weight (kg)	73.5 ± 16.2
BSA at diagnosis (kg/m^2)	1.99 ± 0.24

AL-CA: immunoglobulin light chain cardiac amyloidosis; ATTR: Transthyretin cardiac amyloidosis; ATTR-wt: wild type - transthyretin cardiac amyloidosis; SD: standard deviation; BSA: body surface area

Figure legends

Figure 1. AI-algorithm training workflow.
AI: artificial-intelligence.

Figure 2. Illustrative example comparing LV EF and LVGLS analysis performed with manual tracings and AI derived measurements.
LV EF: left ventricular ejection fraction, LVGLS: left ventricular global longitudinal strain, AI: artificial-intelligence.

Figure 3. Bland-Altman analysis showing the agreement between manual and AI-derived LVEF measurements at A) early timepoint and B) diagnosis timepoint.
AI: artificial-intelligence, LVEF: left ventricular ejection fraction

Figure 4. Correlation analysis between manual and AI LVEF measurements at early timepoint and diagnosis timepoint.
AI: artificial-intelligence, LVEF: left ventricular ejection fraction, r= Pearson Correlation Coefficient

Figure 5. Bland-Altman analysis showing the agreement between manual and AI-derived LV GLS measurements at A) early timepoint and B) diagnosis timepoint.
AI: artificial-intelligence, LVGLS: left ventricular global longitudinal strain

Figure 6. Correlation analysis between manual and AI LVGLS measurements at early timepoint and diagnosis timepoint.
AI: artificial-intelligence, LVGLS: left ventricular global longitudinal strain, r= Pearson Correlation Coefficient