

Assessing automated CMR contouring algorithms using systematic contour quality scoring analysis

Omer M.¹; Amir-Khalili A.¹; Sojoudi A.¹; Thao Le T.²; A Cook S.²; Faye Toh D.²; Bryant J.²; Chin C.²; Miguel Paiva J.¹; Fung K.³; Aung N.³; Y Khanji M.³; Raused E.³; Cooper J.³; E Petersen S.³

¹Circle Cardiovascular Imaging Inc., Calgary, Canada

²National Heart Centre Singapore, Singapore, Singapore

³Barts Heart Centre, London, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: Type of funding sources: Public grant(s) – National budget only. Main funding source(s): SmartHeart EPSRC programme grant (www.nihr.ac.uk), London Medical Imaging and AI Centre for Value-Based Healthcare

Background: Quality measures for machine learning algorithms include clinical measures such as end-diastolic (ED) and end-systolic (ES) volume, volumetric overlaps such as Dice similarity coefficient and surface distances such as Hausdorff distance. These measures capture differences between manually drawn and automated contours but fail to capture the trust of a clinician to an automatically generated contour.

Purpose: We propose to directly capture clinicians' trust in a systematic way. We display manual and automated contours sequentially in random order and ask the clinicians to score the contour quality. We then perform statistical analysis for both sources of contours and stratify results based on contour type.

Data: The data selected for this experiment came from the National Health Center Singapore. It constitutes CMR scans from 313 patients with diverse pathologies including: healthy, dilated cardiomyopathy (DCM), hypertension (HTN), hypertrophic cardiomyopathy (HCM), ischemic heart disease (IHD), left ventricular non-compaction (LVNC), and myocarditis. Each study contains a short axis (SAX) stack, with ED and ES phases manually annotated. Automated contours are generated for each SAX image for which manual annotation is available. For this, a machine learning algorithm trained at Circle Cardiovascular Imaging Inc. is applied and the resulting predictions are saved to be displayed in the contour quality scoring (CQS) application.

Methods: The CQS application displays manual and automated contours in a random order and presents the user an option to assign a contour quality score: 1: Unacceptable, 2: Bad, 3: Fair, 4: Good. The UK Biobank standard operating procedure is used for assessing the quality of the contoured images. Quality scores are assigned based on how the contour affects clinical outcomes. However, as images are presented independent of spatiotemporal context, contour quality is assessed based on how well the area of the delineated structure is approximated. Consequently, small contours and small deviations are rarely assigned a quality score of less than 2, as they are not clinically relevant. Special attention is given to the RV-endo contours as often, mostly in basal images, two separate contours appear. In such cases, a score of 3 is given if the two disjoint contours sufficiently encompass the underlying anatomy; otherwise they are scored as 2 or 1.

Results: A total of 50991 quality scores (24208 manual and 26783 automated) are generated by five expert raters. The mean score for all manual and automated contours are 3.77 ± 0.48 and 3.77 ± 0.52 , respectively. The breakdown of mean quality scores by contour type is included in Fig. 1a while the distribution of quality scores for various raters are shown in Fig. 1b.

Conclusion: We proposed a method of comparing the quality of manual versus automated contouring methods. Results suggest similar statistics in quality scores for both sources of contours.

Abstract Figure 1

