



HHS Public Access

Author manuscript

JACC Cardiovasc Imaging. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

JACC Cardiovasc Imaging. 2018 December ; 11(12): 1917–1918. doi:10.1016/j.jcmg.2018.04.030.

Automated Cardiac MR Scar Quantification in Hypertrophic Cardiomyopathy Using Deep Convolutional Neural Networks

Ahmed S Fahmy, PhD^a, Johannes Rausch, MS^{a,b}, Ulf Neisius, MD, PhD^a, Raymond H Chan, MD^c, Martin S Maron, MD^d, Evan Appelbaum, MD^a, Bjoern Menze, PhD^b, and Reza Nezafat, PhD^a

^aDepartment of Medicine, Cardiovascular Division, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

^bDepartment of Computer Science, Technical University of Munich, Munich, Germany

^cToronto General Hospital, University Health Network, Toronto, Canada

^dHypertrophic Cardiomyopathy Center, Division of Cardiology, Tufts Medical Center, Boston, MA, USA

Scar volume quantified by cardiovascular magnetic resonance (CMR) with late gadolinium enhancement (LGE) is a novel imaging biomarker for risk stratification in patients with hypertrophic cardiomyopathy (HCM) (1). In current practice, scar quantification often relies on manual delineation of the myocardium borders and the hyper-enhanced regions on LGE images, which is subjective, laborious and time-consuming. Variations among different CMR centers and core laboratories reduce the reproducibility of scar quantification (2). Additionally, variable gadolinium kinetics and the patchy multi-foci appearance of hyperenhancements in HCM patients are major challenges to automatic quantification techniques compared to other CMR applications (2).

In this study, we present an initial proof-of-concept for using deep convolutional neural networks (DCN) (5) to automatically quantify left ventricle (LV) mass and scar volume on LGE in HCM patients. We used a U-net DCN architecture with 150 operational layers including batch-normalization, convolutional, rectified-linear, and dropout layers (3). The network parameters were initially set to random values sampled from standard Gaussian distribution. LGE images of 1041 HCM patients (7775 images) acquired and manually segmented by an expert reader as part of a multi-center/multi-vendor study (1) were split into training (80%) and testing (20%) subsets. Stratified random sampling was used to

Correspondence to: Reza Nezafat, Ph.D., Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA, 02215, Phone: 617-667-1747; Fax: 617-975-5480, rnezafat@bidmc.harvard.edu.

Authors email addresses: Ahmed S Fahmy: asfahmy@bidmc.harvard.edu, Johannes Rausch: johannes.rausch@inf.ethz.ch, Ulf Neisius: uneisius@bidmc.harvard.edu, Raymond H Chan: raymond_chan@alumni.utoronto.ca, Martin S Maron: mmaron@tuftsmedicalcenter.org, Evan Appelbaum: eappelbaum@bidmc.harvard.edu, Bjoern Menze: bjoern.menze@tum.de, Reza Nezafat: rnezafat@bidmc.harvard.edu

Relation to Industry: Nothing to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

achieve a balanced number of cases with scar and from each vendor. Data augmentation with elastic deformation, translation, and mirroring of the training images was used to synthetically increase the dataset size and incorporate prior knowledge (e.g. image segmentation should be invariable to image translation). After off-line training phase, the DCN was used to automatically segment the scar and LV in the testing dataset. The segmented scar volume and LV mass were assessed relative to the manual segmentation on per-slice and per-patient bases. The spatial overlap between the automatically and manually segmented regions was assessed by the Dice similarity coefficient (DSC). The testing dataset was assessed by a second expert to grade the image quality (low, medium, and high), and the DCN performance was compared between the different quality levels. To evaluate the performance of scar segmentation for different training/testing subsets, we repeated dataset splitting, training, and testing of our DCN four times and compared the performance of each evaluation.

Automatic segmentation was achieved in 0.26 sec/image (Figure 1A). The DCN detected scar in 52 patients with scar volume $6.1 \pm 7.4 \text{ cm}^3$ compared to manual detection of 60 patients with volume $7.5 \pm 8.5 \text{ cm}^3$. The automatically and manually segmented scar volumes (over all testing images) were strongly correlated in per-patient ($r_s = 0.84$, $r = 0.9$; $p < 0.001$) (Figure 1B) and per-slice ($r_s = 0.81$, $r = 0.84$; $p < 0.001$) analyses. A strong correlation was also observed between the manually and automatically estimated LV mass in per-patient ($r_s = 0.95$, $r = 0.96$; $p < 0.001$) and per-slice ($r_s = 0.93$, $r = 0.93$; $p < 0.001$) analyses. The segmentation accuracy (measured by DSC) between automatic and manual segmentations was 0.57 ± 0.23 (per-patient) and 0.58 ± 0.28 (per-slice) for the scar, and 0.82 ± 0.08 (per-patient) and 0.81 ± 0.11 (per-slice) for the LV. The DSC of LV segmentation was lower in the apical slices compared to other slice locations (0.70 ± 0.2 vs. 0.83 ± 0.1 ; $p < 0.001$). No significant differences were observed in the scar DSC among slices with different image quality levels in per-patient ($p = 0.86$) or per-slice ($p = 0.65$) analyses. Repeating the training/testing of the DCN showed no significant effect on scar DSC in per-patient ($p = 0.64$) or per-slice ($p = 0.23$) analyses. The results of this study show the potential of deep learning as a tool for automatic segmentation of the LV and scar volume in HCM patients, with strong agreement between the automatic and manual segmentations. Limitations of the study include lack of testing using an independent dataset and no inter-observer evaluation of the manual segmentation. Further improvements in the network architecture as well as increasing the training dataset size and diversity are needed to improve the relatively low DSC segmentation score.

Acknowledgments

Funding: The project described was supported in part by National Institutes of Health 1R01HL129185 (Bethesda, MD, USA), 1R21HL127650 (Bethesda, MD, USA); American Heart Association (AHA) 15EIA22710040 (Dallas, TX, USA).

References:

1. Chan RH, Maron BJ, Olivotto I, et al. Prognostic Value of Quantitative ContrastEnhanced Cardiovascular Magnetic Resonance for the Evaluation of Sudden Death Risk in Patients With Hypertrophic Cardiomyopathy. *Circulation* 2014;130:484–95. [PubMed: 25092278]

2. Engblom H, Tufvesson J, Jablonowski R, et al. A new automatic algorithm for quantification of myocardial infarction imaged by late gadolinium enhancement cardiovascular magnetic resonance: experimental validation and comparison to expert delineations in multi-center, multi-vendor patient data. *J Cardiovasc Magn Reson* 2016;18:27. [PubMed: 27145749]
3. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, Proceedings, Part III Cham: Springer International Publishing; 2015 p. 234–41.*

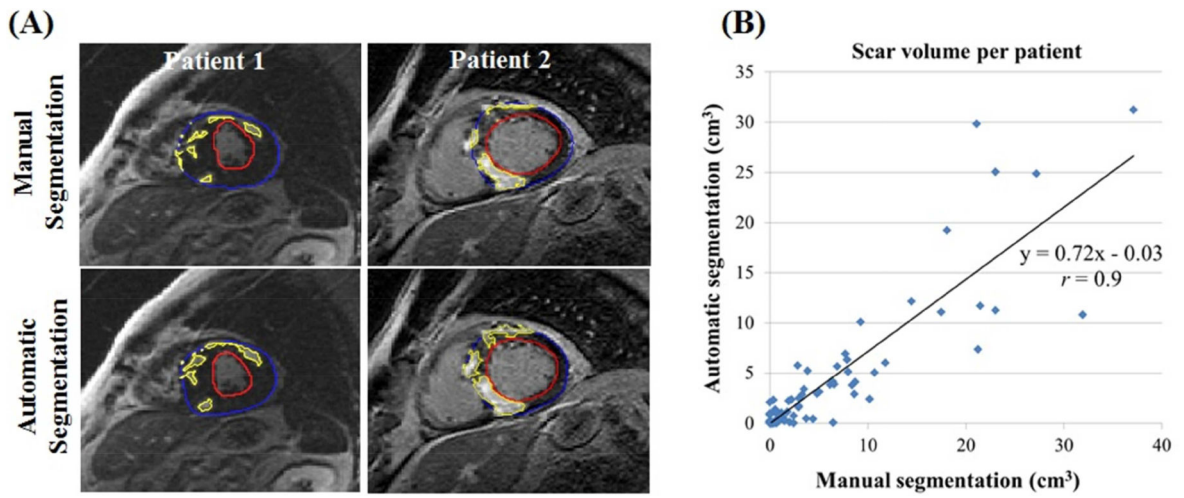


Figure 1. Automatic versus manual scar segmentation.

(A) Scar and myocardium segmentation results for slices from two patients. Contours resulting from manual and automatic segmentations for the epicardium (blue), endocardium (red), and scar (yellow) boundaries are overlaid on late gadolinium enhancement images. (B) Scatter plot of the automatic versus manual segmentation of scar volume.