Multi-Center Variability: Precision Image Analysis compared with Leading Core Laboratories and Institutions for LV Function Quantification

Background

Cardiovascular magnetic resonance (CMR) is the current gold standard for quantifying cardiac ventricular function. Multiple studies have validated the consistency, accuracy, and importance of magnetic resonance (MR) as an imaging modality in the determination of cardiac ventricular chamber size and function [1]. However, the accuracy and reproducibility of 3D segmentation is strongly reader dependent as the isolation of the relevant cardiac anatomy involves a significant degree of manual delineation and reader interpretation of the images [2]. Standard post-processing guidelines have been published by the Society for Cardiovascular Magnetic Resonance (SCMR) to limit reader variability and establish better consistency [3], but these guidelines are not enforced in all institutions and facilities. Variances in clinical training and teaching can impose various nuances of interpretation (such as exclusion of outflow tract) especially across institutions as there is a tendency of reliance on tribal knowledge in lieu of strictly written rule-based protocols concerning CMR post-processing.

In order to quantitatively assess multi-center variability and establish a consensus ground truth dataset, Professor Nagel et al. and the Auckland MRI Research Group collaborated internationally with 7 expert readers from independent CMR core laboratories to analyze a set of CMR cases [4]:

David Bluemke	National Institute of Biomedical Imaging and Bioengineering, USA
Matthias Friedrich	McGill University, Canada
Christopher Kramer	University of Virginia, USA
Raymond Kwong	Harvard Medical School, USA
Sven Plein	University of Leeds, UK
Jeanette Schulz-Menger	Charité University, Germany
Jos Westenberg	Leiden University Medical Center, The Netherlands

Table 1. Contributing expert readers and representative institutions [4].

Image sets were acquired from various scanner vendors with patients exhibiting a variety of pathologies, gender, and age. Each expert independently contoured the complete dataset with their routine software solution of choice, including the papillary muscles and trabeculations in the blood pool, and submitted left ventricular (LV) function results without any prior consensus training.

The quantitative results showcase the multi-center variability among this representative group of internationally recognized experts. Contour data from each reader was evaluated to form a benchmark consensus contour set using the STAPLE algorithm [5] and designated as the ground truth dataset to be used as a standard of reference for CMR image analysis [4].

Readers can download the identical dataset and perform similar LV analyses to compare their own results to the ground truth data set through the Cardiac Atlas Project (www.cardiacatlas.org) [6]. Contours and quantitative data are statistically analyzed to determine the reader's bias, consistency, and variability to the ground truth data set and also to each individual expert.

Multi-Center Variability

Precision Image Analysis (PIA) recognizes the need and importance of standardization and proactively submitted blinded results to the Auckland MRI Research Group to determine how PIA's post-processing protocols perform in comparison to the ground truth dataset and to other participating core laboratories. While each clinical analyst at PIA is extensively trained and internally certified both qualitatively and quantitatively, it is important for every core laboratory to measure consistency to an established industry standard. Clinical analysts at PIA must pass strict internal criteria of limited variability and high reproducibility. In order to do so, each analyst must demonstrate flawless SCMR post-processing protocol practices ensuring that PIA provides the best, highest standard, and most consistent quality of service available.

A single set of unbiased, manually-contoured results from a PIA analyst was submitted to the Auckland MRI Research Group for quantitative assessment to the ground truth benchmark. The PIA representative independently contoured the dataset without any collaboration, over-reads, or previous knowledge of the results in publications or related work on the dataset. The results were compared to the consensus data and indicate that the PIA representative has a small positive bias overall:

EDV = 7.4 ± 5.7 ml
ESV = 2.9 ± 10.4 ml
EF = 0.1 ± 4.3 %
LVM = 5.5 ± 11.3 g

In comparison to the leading expert readers in the field and other core laboratories, the PIA analyst (R8) falls very close to the consensus data benchmark and well within agreement to the expert readers and against leading institutions (Figure 1). Most impressively, not only does PIA's submitted results compare favorably against expert readers in consistency and accuracy (and closer in bias to the consensus benchmark in most instances), PIA's results did not contribute to the formation of the ground truth benchmark, which cannot be said of the expert readers' results. Factoring in this "handicap" reveals that PIA's results are even more impressive than they seem at a first-glance comparison.

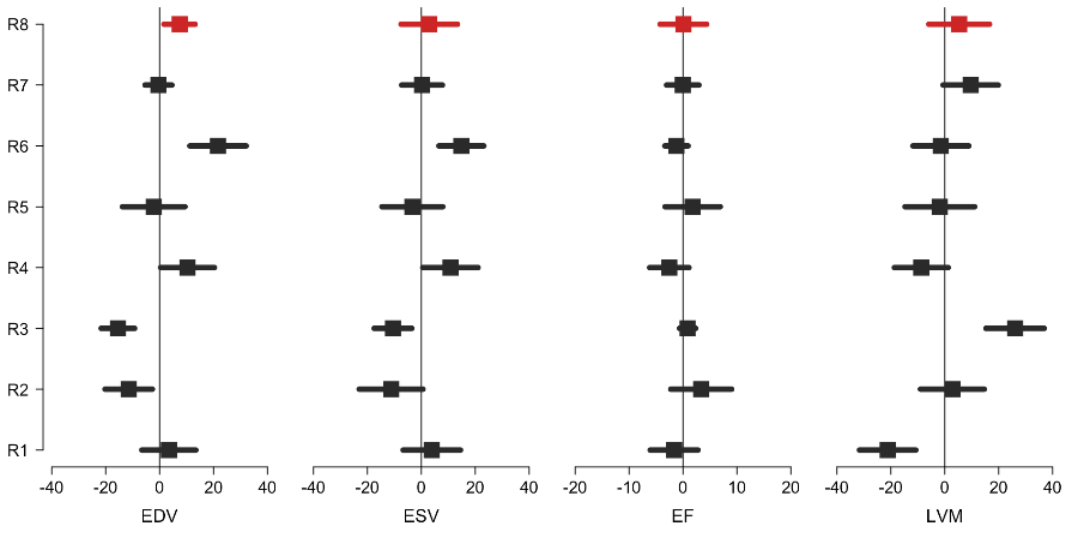


Figure 1. PIA analyst reader (R8) bias and precision against the Ground Truth benchmark (centerline) and expert readers (R1-R7).

LV EF		**LV ESV**	
N:	15	N:	15
Observer Bias:	7.4 ml	Observer bias:	2.9 ml
Standard Dev:	5.7 ml	Standard Dev:	10.4 ml
95 % CI / LOA:	-3.9 to 18.7 ml	95 % CI / LOA:	-17.5 to 23.3 ml
Corr Coefficient r:	0.9967	Corr Coefficient r:	0.9912
95% CI for r:	0.9898 to 0.9989	95% CI for r:	0.9728 to 0.9971
r^2:	0.9934	r^2:	0.9824
LV EF		**LVM**	
N:	15	N:	15
Observer Bias:	0.1 %	Observer Bias:	5.5 g
Standard Dev:	4.3 %	Standard Dev:	11.3 g
95 % CI / LOA:	-8.3 to 8.5 %	95 % CI - LOA:	-16.8 to 27.7 g
Corr Coefficient r:	0.9599	Corr Coefficient r:	0.9432
95% CI for r:	0.8806 to 0.9869	95% CI for r:	0.8338 to 0.9813
r^2:	0.9214	r^2:	0.8896

Table 2. Statistics demonstrating PIA's high accuracy and correlation against the consensus ground truth benchmark.

Visual Assessment and Contour Quality

The Auckland MRI Research Group noted all discrepancies in each of the 318 contoured images. PIA's mean distance, max distance, and the standard deviation of displacement from the consensus contours were reported for each image in detail (Figure 2). PIA further analyzed the data by normalizing and subdividing the cardiac scans equally into thirds by slice location to represent the apical, mid, and basal regions. The basal and apical regions demonstrated the largest deviations in 2D contour displacement regardless of phase, due to higher variations of interpretation of the more complex anatomical morphology of the outflow tract insertion and atrioventricular boundaries at the base, and inherent volume averaging at the apex. PIA's apical LV endocardium contours in end diastole (ED) varied an overall an average of 1.10 ± 0.98 mm from the consensus contours. The mid-ventricular slices were the most consistent and accurate at an average of 0.54 ± 0.27 mm. PIA's basal slices were an average of 0.76 ± 0.48 mm from the consensus benchmark (Figure 3).

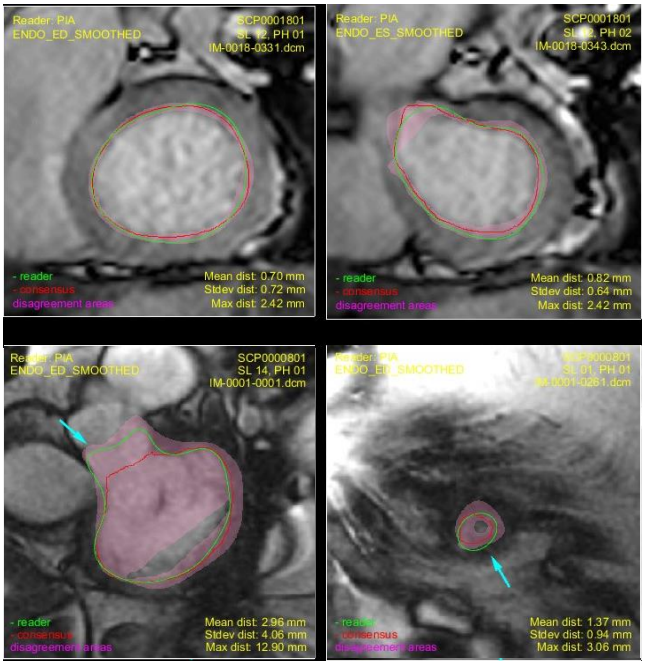


Figure 2. Sampling of contour grading results. GREEN: PIA contours, RED: Consensus, PINK: Areas of expert reader disagreements. ARROW: >2 stdev difference from Consensus.

To obtain global LV statistics for each case, the mean and standard deviations were averaged with every image equally weighted (Appendix A). Global mean distance from the consensus contours were no more than 1.42 mm for all cases averaging 0.85 ± 0.84 mm overall for the data set. The maximum difference in contour discrepancy was reported to be 13.53 mm in the heart failure case due to a consistent difference in interpretation of the LV outflow tract and aortic valve cutoff boundary (Figure 2 – lower left).

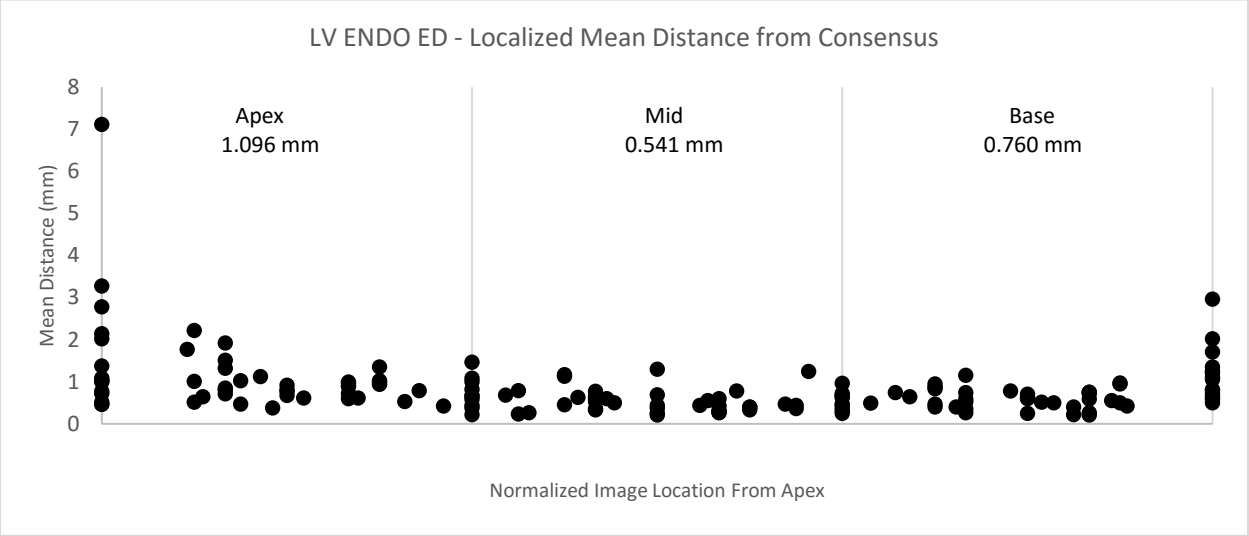


Figure 3. Scatterplot displaying PIA's mean distance from the ground-truth consensus contours on a slice-by-slice basis for LV endo in ED phase for all 15 cases. Regional average difference is displayed.

Summary

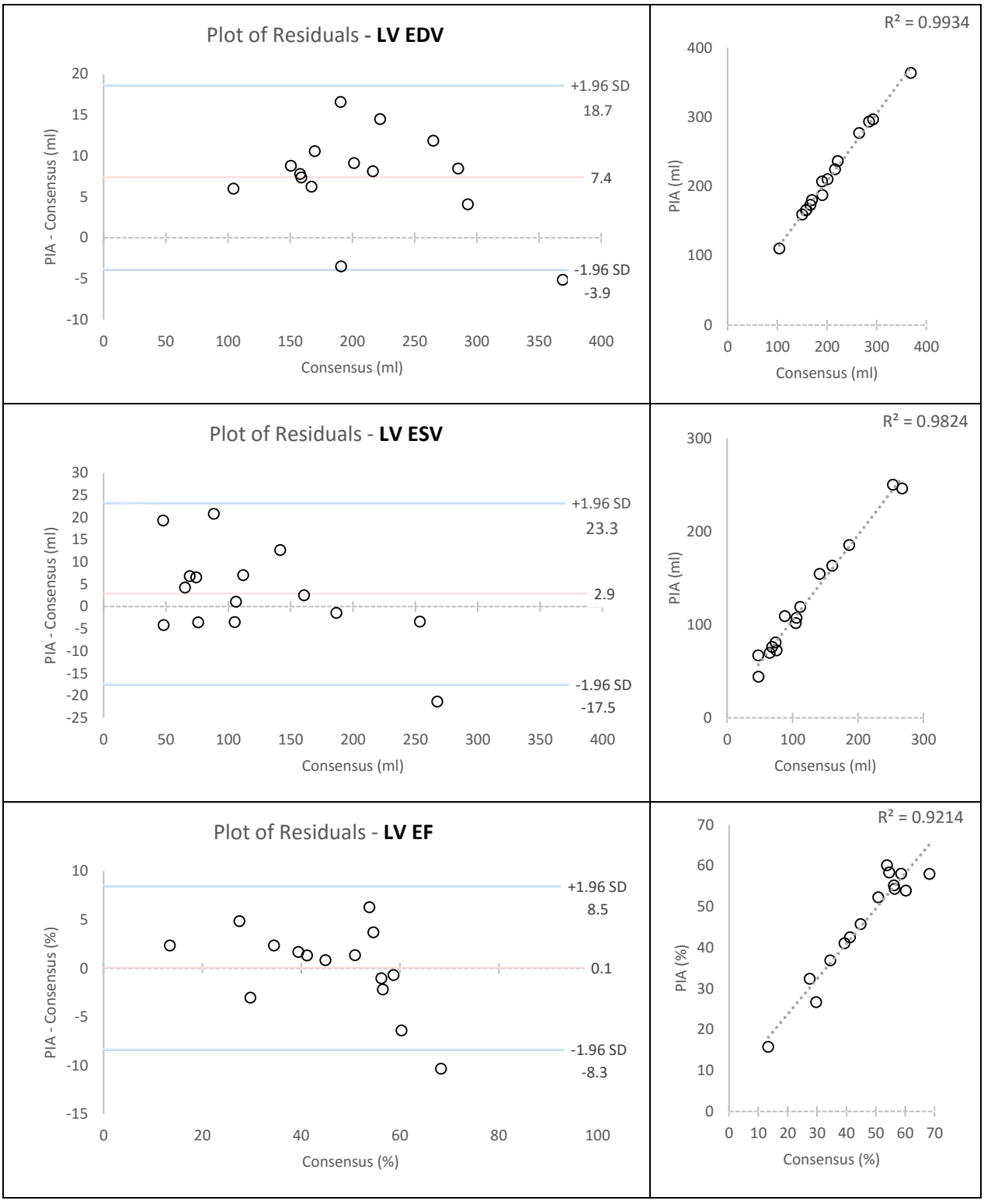
Thorough and detailed analyses of both the quantitative volumetric differences and absolute contour displacement of PIA's results demonstrate that PIA's internal standard operating procedures and PIA's interpretation and execution of SCMR protocols correlate strongly with the consensus benchmark and against a team of expert readers. On overall closeness (bias and standard deviation) to the consensus benchmark, PIA impressively ranks 2rd out of 8 against renowned experts even when handicapped - as PIA's results were the only ones that did not contribute to the formation of the benchmark consensus results. Correlation coefficients (r) of PIA's LV function report deliverables against the benchmark range from 0.943 to 0.997 with a slight positive bias. The quantitative results showcase PIA's impressive conformance to published international standards and is a robust testament to the high quality and attention-to-detail that rivals international experts in the field. PIA strives for continual improvements in training, performance, and internal post-processing standards and will use this information to progress and further that goal in order to provide the highest quality of service possible. Quality, consistency, reproducibility, and standardization are cornerstones that all PIA analysts strive to embody.

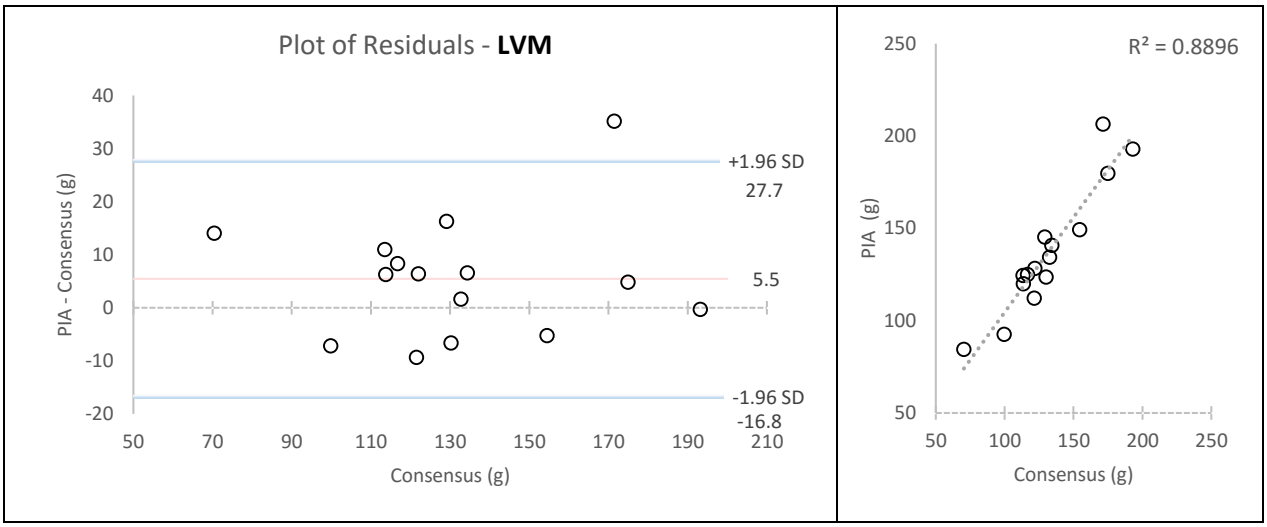
APPENDIX A - Tabulated Global Contour Discrepancies (mm)

	ENDO						EPI		
	ED			**ES**			**ED**		
Pathology	Mean Dist	Stdev Dist	Max Dist	Mean Dist	Stdev Dist	Max Dist	Mean Dist	Stdev Dist	Max Dist
Healthy	0.54	0.73	3.98	0.74	0.74	3.76	0.99	0.83	3.98
Healthy	0.82	0.76	4.44	1.04	0.80	4.44	0.78	0.74	3.89
Healthy	0.71	0.73	2.89	1.14	0.91	5.78	0.77	0.80	3.23
Healthy	1.42	0.79	9.22	0.69	0.77	3.22	0.71	0.82	4.07
Healthy	1.11	0.80	5.14	1.17	0.84	4.37	1.12	1.13	6.06
Heart Failure	0.57	0.60	2.71	0.56	0.65	3.63	0.57	0.63	3.83
Heart Failure	0.79	0.97	12.90	0.77	0.79	4.10	1.20	1.08	13.53
Hypertrophy	0.69	0.76	4.10	1.33	1.13	4.93	0.62	0.71	2.73
Hypertrophy	0.76	0.68	4.31	0.85	0.80	4.31	1.09	0.87	6.09
Infarct	0.77	0.95	4.66	0.68	0.97	4.17	0.76	0.96	2.95
Infarct	0.91	0.86	5.97	0.82	0.86	6.29	1.06	0.85	5.07
Infarct	0.50	0.83	4.19	0.57	0.86	4.19	0.72	0.94	4.19
Infarct	0.80	0.88	4.45	0.71	0.79	4.45	0.74	0.88	4.22
Infarct	0.94	0.92	5.80	0.74	0.84	7.03	0.87	0.92	7.57
Infarct	0.82	0.77	5.71	0.90	0.88	4.12	1.18	0.86	5.11

PIA's globally averaged discrepancies from the ground truth contours in millimeters (mm). RED values indicate the largest differences in each category. For mean and standard deviation of the distance, each image was weighted equally to obtain the global values.

APPENDIX B - Calibration Plots





Bland-Altman calibration plots and statistical data of PIA's results to the published ground truth benchmark results.

References

[1] W. Hundley, D. Bluemke, J. Finn, S. Flamm, M. Fogel, F. MG, V. B. Ho, M. Jerosch-Herold, C. M. Kramer, W. J. Manning, M. Patel, G. M. Pohost, A. E. Stillman, R. D. White and P. K. Woodard, "ACCF/ACR/AHA/NASCI/SCMR 2010 Expert Consensus Document on Cardiovascular Magnetic Resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents.," *Circulation,* pp. 462-508, 2010.

[2] T. Karamitsos, L. Hudsmith, J. Selvanayagam, S. Neubauer and J. Francis, "Operator induced variability in left ventricular measurements with cardiovascular magnetic resonance is improved after training," *Journal of Cardiovascular Magnetic Resonance,* pp. 777-783, 2007.

[3] J. Schulz-Menger, D. A. Bluemke, J. Bremerich, S. D. Flamm, M. A. Fogel, M. G. Friedrich, R. J. Kim, F. v. Knobelsdorff-Brenkenhoff, C. M. Kramer, D. J. Pennell, S. Plein and E. Nagel, "Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) Board of Trustees Task Force on Standardized Post Processing," *Journal of Cardiovascular Magnetic Resonance,* pp. 15-35, 2013.

[4] A. Suinesiaputra, D. A. . Bluemke, B. R. Cowan, M. G. Friedrich and C. M. Kramer, "Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours," *Journal of Cardiovascular Magnetic Resonance,* 2015.

[5] S. Warfield, "Computational Radiology Laboratory," 2015. [Online]. Available: http://crl.med.harvard.edu/.

[6] A. Suinesiaputra, B. R. Cowan, A. O. Al-Agamy, M. A. Elattar, N. Ayache, A. S. Fahmy, A. M. Khalifa, P. Medrano-Gracia, M.-P. Jolly, A. H. Kadish, D. C. Lee, J. Margeta, S. K. Warfield and A. A. Young, "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images," *Medical Image Analysis,* pp. 50-62, 2014.