


RESEARCH

Open Access



Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification

Alex Bratt¹, Jiwon Kim^{1,2}, Meridith Pollie², Ashley N. Beecy², Nathan H. Tehrani², Noel Codella³, Rocio Perez-Johnston⁴, Maria Chiara Palumbo², Javid Alakbarli², Wayne Colizza¹, Ian R. Drexler¹, Clerio F. Azevedo⁵, Raymond J. Kim⁵, Richard B. Devereux² and Jonathan W. Weinsaft^{1,2,4,6*} 

Abstract

Background: Phase contrast (PC) cardiovascular magnetic resonance (CMR) is widely employed for flow quantification, but analysis typically requires time consuming manual segmentation which can require human correction. Advances in machine learning have markedly improved automated processing, but have yet to be applied to PC-CMR. This study tested a novel machine learning model for fully automated analysis of PC-CMR aortic flow.

Methods: A machine learning model was designed to track aortic valve borders based on neural network approaches. The model was trained in a derivation cohort encompassing 150 patients who underwent clinical PC-CMR then compared to manual and commercially-available automated segmentation in a prospective validation cohort. Further validation testing was performed in an external cohort acquired from a different site/CMR vendor.

Results: Among 190 coronary artery disease patients prospectively undergoing CMR on commercial scanners (84% 1.5T, 16% 3T), machine learning segmentation was uniformly successful, requiring no human intervention: Segmentation time was < 0.01 min/case (1.2 min for entire dataset); manual segmentation required 3.96 ± 0.36 min/case (12.5 h for entire dataset). Correlations between machine learning and manual segmentation-derived flow approached unity ($r = 0.99$, $p < 0.001$). Machine learning yielded smaller absolute differences with manual segmentation than did commercial automation (1.85 ± 1.80 vs. 3.33 ± 3.18 mL, $p < 0.01$): Nearly all (98%) of cases differed by ≤ 5 mL between machine learning and manual methods. Among patients without advanced mitral regurgitation, machine learning correlated well ($r = 0.63$, $p < 0.001$) and yielded small differences with cine-CMR stroke volume ($\Delta 1.3 \pm 17.7$ mL, $p = 0.36$). Among advanced mitral regurgitation patients, machine learning yielded lower stroke volume than did volumetric cine-CMR ($\Delta 12.6 \pm 20.9$ mL, $p = 0.005$), further supporting validity of this method. Among the external validation cohort ($n = 80$) acquired using a different CMR vendor, the algorithm yielded equivalently small differences ($\Delta 1.39 \pm 1.77$ mL, $p = 0.4$) and high correlations ($r = 0.99$, $p < 0.001$) with manual segmentation, including similar results in 20 patients with bicuspid or stenotic aortic valve pathology ($\Delta 1.71 \pm 2.25$ mL, $p = 0.25$).

(Continued on next page)

* Correspondence: jww2001@med.cornell.edu

¹Department of Radiology, Weill Cornell Medicine, 525 E 68th St, New York, NY 10065, USA

²Greenberg Division of Cardiology, Department of Medicine, Weill Cornell Medicine, 525 E 68th St, New York, NY 10065, USA

Full list of author information is available at the end of the article



(Continued from previous page)

Conclusion: Fully automated machine learning PC-CMR segmentation performs robustly for aortic flow quantification - yielding rapid segmentation, small differences with manual segmentation, and identification of differential forward/left ventricular volumetric stroke volume in context of concomitant mitral regurgitation. Findings support use of machine learning for analysis of large scale CMR datasets.

Keywords: Cardiovascular magnetic resonance, Machine learning, Deep learning, Phase contrast, Aorta

Background

Cardiovascular magnetic resonance (CMR) is increasingly employed for assessment of patients with known or suspected valvular heart disease. Phase contrast (PC) imaging is central for this application, as it enables quantification of valvular flow and velocity. When acquired through the aortic annulus, PC-CMR also provides an index of left ventricular (LV) stroke volume – enabling LV systolic performance to be measured independent of volumetric analysis. PC-CMR has been well validated in relation to invasive cardiac performance indices, and shown to provide improved predictive value vs. other non-invasive methods such as echocardiography [1–5]. Given widespread utility of PC-CMR, accurate and widely applicable analytic methods for this pulse sequence are of substantial importance.

In routine clinical practice, PC-CMR analysis requires manual segmentation of multiple images (~20–30) acquired throughout the cardiac cycle. Given that high temporal resolution is critical for accurate flow quantification, PC-CMR typically entails acquisition of more cardiac frames than do other gated CMR pulse sequences (e.g. cine). Additionally, PC-CMR is often acquired in multiple 2D orientations, adding to cumulative number of images needed for segmentation. Automated PC-CMR segmentation methods are commercially available, but border tracking can be suboptimal – requiring time-consuming operator adjustments that compromise practicality of large-scale studies.

New advances in machine learning have improved capabilities for automated image processing. Convolutional neural networks (CNN) have been a key area of focus, and have been shown to yield human or superhuman performance in a variety of medical classification tasks, including detection of diabetic retinopathy [6] and classification of pulmonary fibrosis [7]. CNN-based segmentation (i.e. pixel-wise classification) models have been applied to CMR, with focus on cine-CMR chamber volumes and systolic function [8–14]. However, to date, fully automated machine learning has not been applied for PC-CMR quantification of valvular flow.

This study tested a novel, machine learning derived, fully automated algorithm for aortic flow quantification on PC-CMR – the goal was to test the incremental utility of machine learning derived (fully automated) flow

segmentation to conventional PC-CMR analysis using commercially available automated and manual segmentation. To do so, a broad multicenter cohort of patients undergoing CMR using equipment from different commercial vendors was studied, including patients with coronary artery disease (CAD) as well as an independent cohort enriched for aortic valve pathology.

Methods

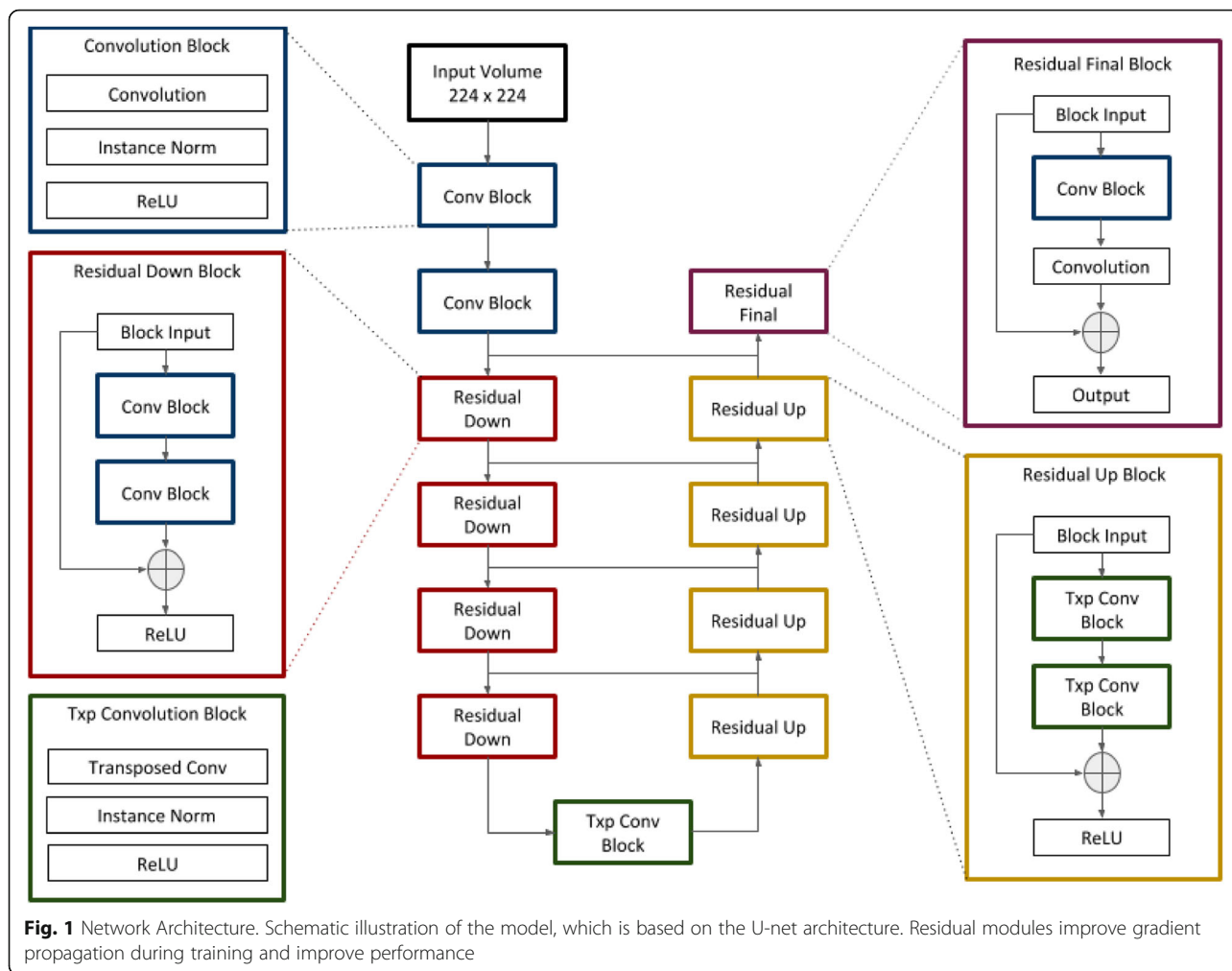
Model and training

The automated segmentation model was based on neural network architecture described by Han [15], a modified U-net [16], for which excellent performance has been previously demonstrated in medical segmentation: The model makes use of residual modules [17], which improve gradient flow between adjacent layers and increase classification accuracy. A diagram of the model's architecture is shown in Fig. 1.

The machine learning algorithm was initially trained in a *derivation cohort*, consisting of consecutive patients ($n = 150$) who underwent clinical CMR (with aortic valve PC-CMR) between January – November 2017. For each exam, manual segmentation maps were generated from PC-CMR: This entailed labeling pixels in the magnitude images as either valve or non-valve using 3DSlicer [18], an open-source medical image post-processing application. Prior to neural network processing, input images were resampled and (if necessary) zero-padded to 256×256 pixels. Pixel intensity values in the magnitude images were then rescaled to values between zero and one. Model training was performed using magnitude images as input and corresponding ground-truth manual segmentation maps as output (Fig. 2). The training set contained a total of 4345 unique images from 150 patients. Aggressive data augmentation was employed at batch time in the form of random zoom, rotation, crop (224×224), horizontal/vertical flip, and addition of Gaussian noise. A weighted softmax/cross entropy loss function was used for training as follows:

$$\text{loss}(x, i) = -w[i] * \ln \frac{e^{x[i]}}{\sum_j^C e^{x[j]}} \quad (1)$$

where x is the output logit vector at a given pixel, i the true class label, w the vector of class weights, and C the number of classes. Weighting was employed to combat



class imbalance given that the vast majority of pixels in each image were non-valve. A class weight of 0.2 was empirically assigned to the non-valve class and 0.8 to the valve class. RMSProp was used to apply incremental parameter updates.

The model was built in Python using the deep learning framework Pytorch. Training and testing were performed on a workstation with four CPU cores, 64 GB of system memory, and a graphics processing unit (GPU) with 11 GB of video memory (NVIDIA [Santa Clara, California, USA] GTX 1080 Ti). Software code pertaining to both training and testing of the machine learning model can be found on line at: https://github.com/akbratt/PC_AutoFlow.

Flow calculation

Data was extracted from PC-CMR DICOM files. Established algorithms were used to convert raw phase map pixel intensities to velocities as follows:

$$v = P * M * ASF \tag{2}$$

such that

$$ASF = \frac{10\pi R}{VENC} \tag{3}$$

where P and M represents raw pixel values from the phase and magnitude maps, respectively, ASF is an amplitude scaling factor, R is a reconstruction scaling factor specified in the DICOM header, and $VENC$ is an adjustable scanner parameter representing the maximum measurable flow velocity.

Flow was calculated from the automated segmentation map of a given phase contrast scan as:

$$NetFlow = \sum_{n=0}^N \sum_{i=0}^I S_{n,i} V_{n,i} a\Delta t \tag{4}$$

where n is the segment index, N the number of temporal segments in the scan, I the number of pixels in each

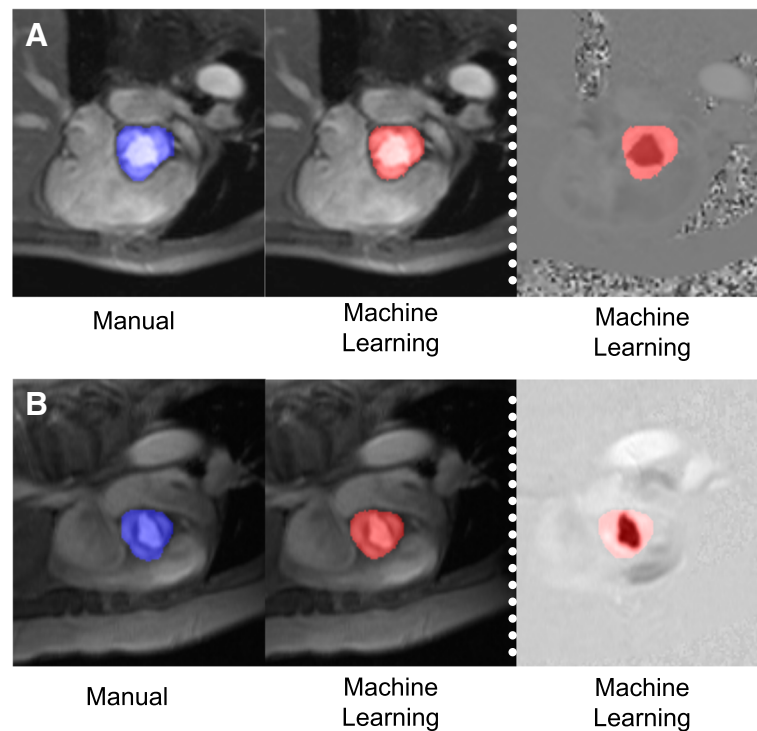


Fig. 2 Representative Examples. Typical examples of aortic contouring as performed by manual segmentation (blue, left) and fully automated machine learning (red, right) in a patient with trileaflet (a) and bicuspid (b) aortic valve. For both examples, magnitude images shown on left, and corresponding PC images shown on right

segment, S the binary segmentation map, V the velocity map calculated using eqs. 1–3, a the pixel area (in cm^2), and Δt the time interval between segments.

Study population

An independent validation cohort was thereafter employed to test the algorithm, which was comprised of CAD patients ($n = 190$) enrolled in two prospective (Cornell) institutional protocols focused on LV remodeling. PC-CMR exams were performed using a standardized protocol, in which PC-CMR datasets were acquired (through plane) at the level of the aortic valve leaflet tips and cine-CMR datasets (for assessment of systolic function) were acquired in contiguous short axis slices (6 mm slice thickness, 4 mm gap) throughout the LV. CMR exams in the validation cohort were performed using a 3T CMR scanner (84% 1.5 T, 16% 3T; General Electric Healthcare, Waukesha, Wisconsin, USA) scanners. Typical PC-CMR parameters were as follows: flip angle = 20 deg., Venc = 150–350 cm/sec, TR [1.5T] = 8 msec, TE [1.5T] = 3.7 msec, TR [3T] = 5 msec, TE = 3.6 msec. Transthoracic echocardiography was performed within one week of CMR (99% within 24h) in accordance with standardized protocol as previously detailed for each of the two prospective studies from which the current cohort was derived [19,

20]. Clinical and demographic information was prospectively acquired at time of study enrollment.

This research protocol was performed with approval of the Weill Cornell Institutional Review Board (IRB), which approved retrospective analysis of pre-existing datasets utilized for model training (derivation cohort). Validation cohort patients provided written informed consent for research participation.

Volume overlap and surface distance metric analysis

The automated segmentation model was evaluated in terms of volume overlap and surface distance metrics by comparing automated segmentation maps to corresponding ground-truth manual segmentation maps. Volume overlap and surface distance metrics were tested on all scans in the derivation cohort ($n = 150$) using six-fold cross-validation. Cross validation is a procedure whereby data is randomly split into non-overlapping subsets such that a model can be trained on all but one subset and tested on the remaining subset. In this case, a different model instance was trained and tested for each of the 6 hold-out subsets and test metrics were averaged per-case for the entire dataset.

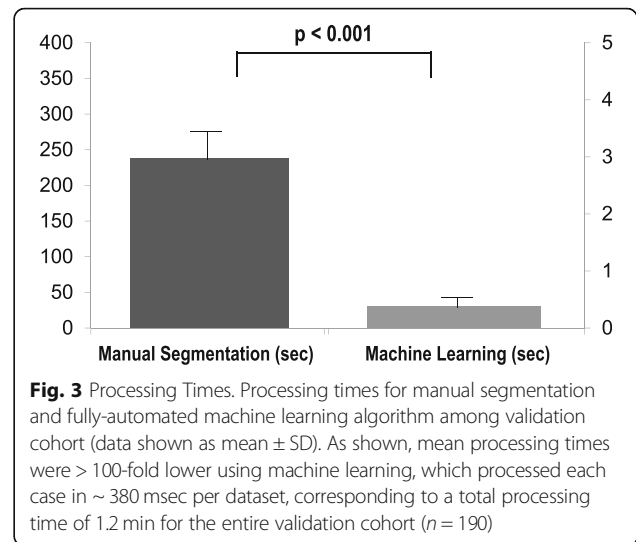
Volume overlap metrics (Dice and Jaccard coefficients) consider only pixels that are labeled as valve. These coefficients take on values between zero and

Table 1 Patient characteristics

	Overall (n = 190)
Clinical	
Age (years)	57 ± 12
Male gender	87% (165)
Body surface area	2.0 ± 0.2
Coronary Artery Disease Risk Factors	
Hypertension	47% (90)
Hypercholesterolemia	54% (102)
Diabetes mellitus	28% (53)
Tobacco use	35% (66)
Family history	30% (56)
Cardiovascular Medications	
Beta-blocker	91% (173)
ACEI/ARB	60% (113)
Loop diuretic	15% (28)
Statin	93% (177)
Aspirin	98% (186)
Thienopyridine	83% (158)
Warfarin	5% (9)
Nitroglycerin	13% (25)
Cardiac morphology/function	
Left Ventricle	
Ejection fraction (%)	52.2 ± 13.3
LV dysfunction (EF <= 55%)	55% (105)
End-diastolic volume (ml)	161.9 ± 49.2
End-systolic volume (ml)	81.6 ± 46.4
Myocardial mass (g)	137.9 ± 38.2
Late gadolinium enhancement (present)	98% (186)
Infarct size (% myocardium)	14.5 ± 10.4
Aortic Valve	
Bileaflet	2% (3)
Thickening/ fibrocalcific changes	12% (23)
Stenosis	2% (4)
Regurgitation	7% (13)

ACEI angiotensin converting enzyme inhibitor, ARB angiotensin receptor blocker

one such that a value of one is perfect overlap between segmentation maps and zero is no overlap. Distance metrics (Hausdorff [HD] and average symmetric surface [ASSD] distances) operate on binary surface plots generated from volumetric segmentations by zeroing any valve pixels with no neighboring non-valve pixels. Equations for volume overlap and surface distance metrics are shown in Additional file 1.



Flow comparisons / algorithm evaluation

Net forward trans-aortic flow calculated via fully automated machine learning segmentation was compared to that generated by manual segmentation, which was performed by an experienced (level III trained) physician (JWW). Flow differences between manual and machine learning were compared to those between manual and a conventional (commercially available) automated algorithm (Cardiac VX, General Electric Healthcare, Waukesha, Wisconsin, USA): The commercial algorithm requires a user to manually contour a single temporal segment; the segmentation mask is then propagated to all other temporal segments with automatic adjustments to account for valve motion and deformation. To directly test incremental utility of the machine learning approach, analyses using the commercial algorithm were in no way adjusted following initial segmentation.

Intra- and inter-reader reproducibility for manual, conventional automated, and machine learning segmentation were determined via analysis of a random subset of 20 patients.

External validation

Data from an additional institution (Duke) was used to further test robustness of the model, representing 130 CMR scans acquired using different vendor (Siemens, Munich, Germany) equipment (53% 3T, 47% 1.5T) in a cohort enriched for patients (n = 40) with clinically documented aortic valve pathology (bicuspid aortic valve [BAV]) or aortic stenosis [AS]; 25% mild / 35% moderate / 40% severe). To do so, a new instance of the model was trained on a dataset consisting of the entire derivation cohort (n = 150) as well as 50 exams from the external validation cohort (including 10 with AS and 10 with BAV; total n = 200). The model was then tested on the remainder of the external validation cohort (n = 80),

including an equivalent number of patients with aortic valve pathology ($n = 10$ AS, $n = 10$ BAV). The Duke IRB provided approval for analysis of de-identified datasets for research purposes.

Statistical methods

Comparisons between groups were made using Student's *t*-test (expressed as mean \pm standard deviation [SD]) for continuous variables. Inter and intra-observer agreement between methods was assessed using the method of Bland and Altman [21], which yielded mean difference as well as limits of agreement between methods (mean \pm 1.96 SD). Bivariate correlation coefficients, intra-class correlation coefficients, and linear regression equations were used to evaluate associations between variables. Statistical calculations were performed using SPSS 24.0 (Statistical Package for the Social Sciences, International Business Machines, Inc., Armonk, New York, USA), SciPy [22], and Excel (Microsoft Inc., Redmond, Washington, USA). Two-sided $p < 0.05$ was considered indicative of statistical significance.

Results

Model training, volume overlap, and surface distance metric analysis

Model training was initially performed in the derivation cohort and deemed complete based on training set Dice coefficient plateau. Volume overlap and surface distance metrics during cross validation demonstrated excellent magnitude of agreement between manual and model generated segmentations (mean Dice = 0.940 [CI 0.937–0.943], Jaccard = 0.888 [CI 0.883–0.893], HD = 3.5 mm [CI 3.1–3.9], ASSD = 0.7 mm [CI 0.6–0.8]).

Clinical application and external validation

Machine learning PC-CMR segmentation was subsequently tested in a cohort of 190 CAD patients undergoing CMR using commercial (GE) scanners, for whom demographics are reported in Table 1. Figure 2 shows representative examples of manual and model-generated data, including cases of tricuspid and bicuspid aortic valve. Segmentation was successful in all cases and required no user intervention.

As shown in Fig. 3, segmentation time was < 0.01 min per case for machine learning segmentation (1.2 min for entire dataset) with GPU acceleration, whereas manual

segmentation required an average time of 3.96 ± 0.36 min per case (12.5 h for entire dataset). Automated per-case segmentation time averaged 19.04 s without GPU acceleration.

Reproducibility

Table 2 reports intra- and inter-observer variability (mean difference in net flow) for manual, conventional automated, and machine learning segmentation methods. As shown, both manual and conventional automated methods yielded good reproducibility, as evidenced by non-significant mean differences and small limits of agreement as well as high intra-class correlation coefficients (ICC) for manual (inter- and intra-rater ICC both > 0.99) and conventional automated (both > 0.97) segmentations. Since the machine learning model is deterministic with respect to input (apart from correctly loading DICOMs for a given dataset), inter- and intra-observer variability for model-generated segmentations was zero and intraclass correlation coefficients were 1.0.

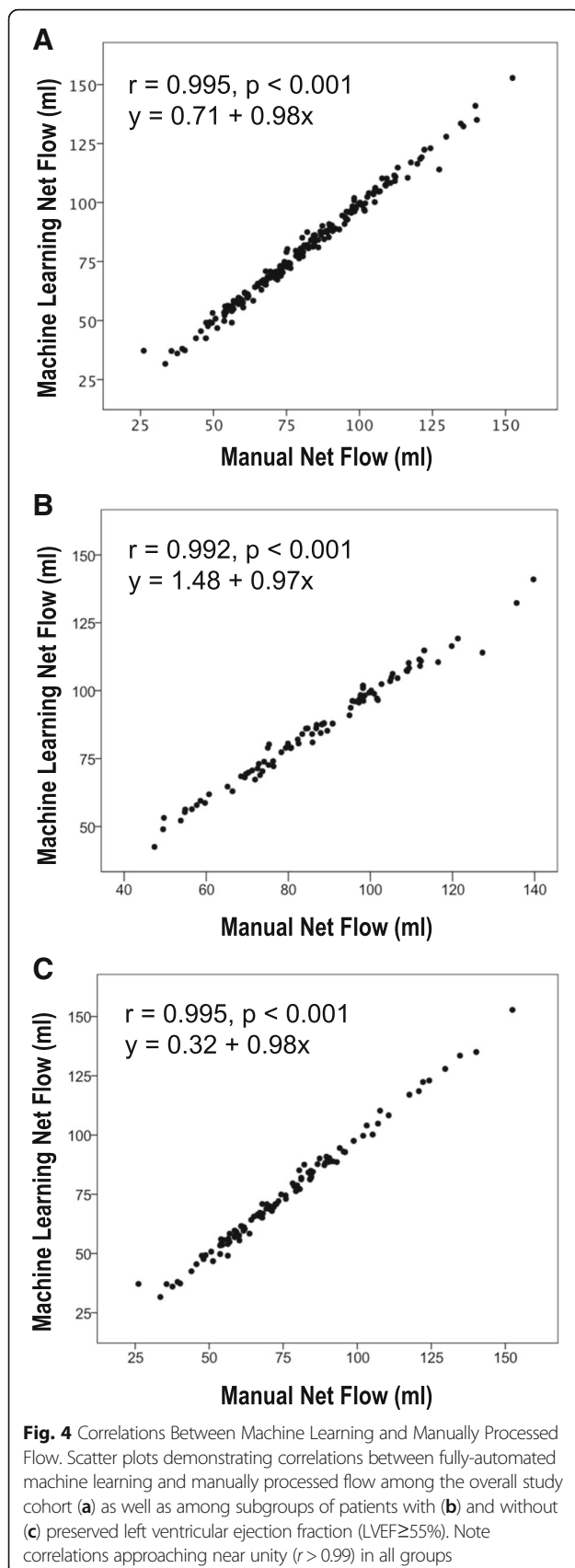
Comparisons to conventional flow segmentation

As illustrated in Fig. 4, correlations between machine learning- and manual segmentation-derived flow approached unity ($r = 0.99$) among the overall study cohort, as well as among patients with ($n = 102$) and without ($n = 88$) LV systolic dysfunction (LV ejection fraction [EF] $< 55\%$). Table 3 details flow quantification results by respective segmentation methods. As shown, both machine learning and conventional automated methods yielded extremely small, albeit statistically significant absolute differences vs. manual segmentation, although magnitude of difference was smaller for machine learning vs. conventional automation (1.85 ± 1.80 vs. 3.33 ± 3.18 mL, $p < 0.01$). Machine learning also yielded a slightly higher intra-class correlation coefficient in relation to manual segmentation (ICC = 0.994) than did conventional automated segmentation (ICC = 0.980).

Regarding methodological differences, Fig. 5 reports Bland-Altman plots for machine learning data in relation to manual segmentation: As shown, 98% ($n = 186$) of cases differences were ≤ 5 mL: In 2 cases with marked discrepancies (~ 10 mL), prominent artifact (1 prominent peri-aortic dephasing artifact, 1 motion artifact) may have been responsible. Notably, in the two cases for which machine learning yielded greatest difference with manual

Table 2 Reproducibility Analyses

	Intra-Observer			Inter-Observer		
	Mean \pm SD (mL)	Limits of Agreement (mL)	p	Mean \pm SD (mL)	Limits of Agreement (mL)	p
Manual	0.18 \pm 1.6	-3.0 to 2.5	0.32	-0.28 \pm 1.2	-2.6 to 2.1	0.62
Conventional	-0.33 \pm 1.8	-3.8 \pm 3.1	0.42	0.28 \pm 3.0	- 5.7 \pm 6.3	0.69
Machine Learning	0	0	-	0	0	-



segmentation, conventional automated and manual methods also yielded a substantial difference in one case (12.5 mL) and a lesser difference in the other (3.6 mL).

Performance in relation to LV dysfunction and mitral regurgitation

Cine-CMR stroke volume was employed as an independent means of assessing machine learning derived aortic flow; echocardiography was used as an independent arbiter for presence/severity of mitral regurgitation. As shown in Fig. 6a, correlations between the two approaches were good ($r = 0.63, p < 0.001$) among patients without advanced (>mild) mitral regurgitation (a known cause for differential volumetric and forward LV stroke volume) – paralleling small, non-significant differences in mean stroke volume ($\Delta 1.3 \pm 17.7$ mL, $p = 0.36$). Manual and conventional automated methods yielded similar magnitude of correlation with LV volumetric stroke volume ($r = 0.65, p < 0.001$ for both), paralleling non-significant differences in LV stroke volume as calculated by cine-CMR volumetric segmentation and respective PC-CMR segmentation methods (both $p = \text{NS}$).

As shown in Fig. 6b, machine learning-derived segmentations yielded lower stroke volume than did volumetric cine-CMR data among patients with advanced (>mild) mitral regurgitation ($\Delta 12.6 \pm 20.9$ mL, $p = 0.005$), consistent with decreased forward cardiac flow in context of atrio-ventricular valvular insufficiency.

External validation

Further evaluation of the machine learning algorithm in the external (Siemens) dataset demonstrated equivalent model performance. There was excellent agreement between manual and automated segmentation in the external validation cohort with respect to volume overlap (mean Dice = 0.940 [CI 0.936–0.944], mean Jaccard = 0.887 [0.881–0.893]). Similarly, manual and machine learning derived flow measurements yielded small mean difference in net flow ($\Delta 1.39 \pm 1.77$ mL, $p = 0.8$) and high magnitude of correlation ($r > 0.99, p < 0.001$). Among the subgroup of 20 patients with aortic valve pathology encompassed in the test set of the validation cohort ($n = 10$ BAV, $n = 10$ AS), machine learning yielded similarly small mean differences ($\Delta 1.71 \pm 2.25$ mL, $p = 0.25$) and high correlations ($r > 0.99, p < 0.001$) with manual segmentation.

Discussion

This study—testing a novel application of machine learning for fully automated aortic flow quantification—demonstrates several key findings: First, machine learning performed robustly among a broad cohort of CAD patients, as evidenced by successful segmentation in all cases and rapid processing time ($n = 190$ cases,

Table 3 Difference in net flow between manual segmentation in relation to machine learning and conventional (commercially available) automated segmentation

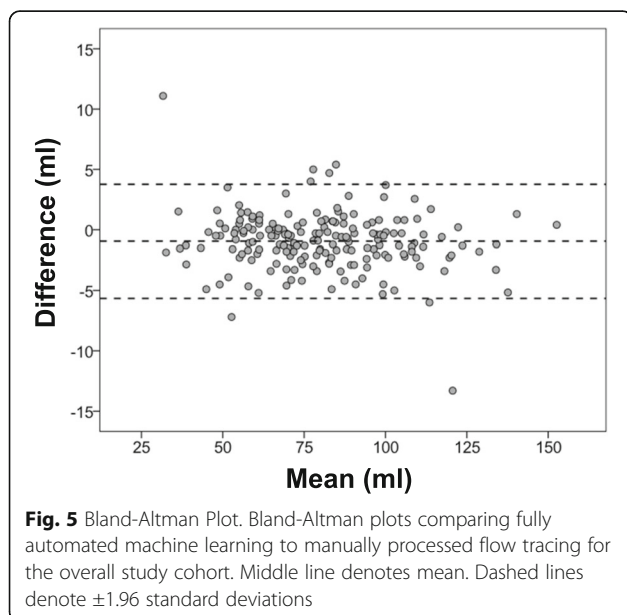
	Net Flow	Absolute Difference (manual – method)	Correlation
Manual	81.5 ± 24.2 mL		
Machine Learning	80.5 ± 23.7 mL ^a	1.85 ± 1.80 mL ^b	$y = 1.01x + 0.16$ $r^2 = 0.99, p < 0.001$
Conventional	80.1 ± 23.2 mL ^a	3.33 ± 3.18 mL ^b	$y = 1.02x - 0.31$ $r^2 = 0.96, p < 0.001$

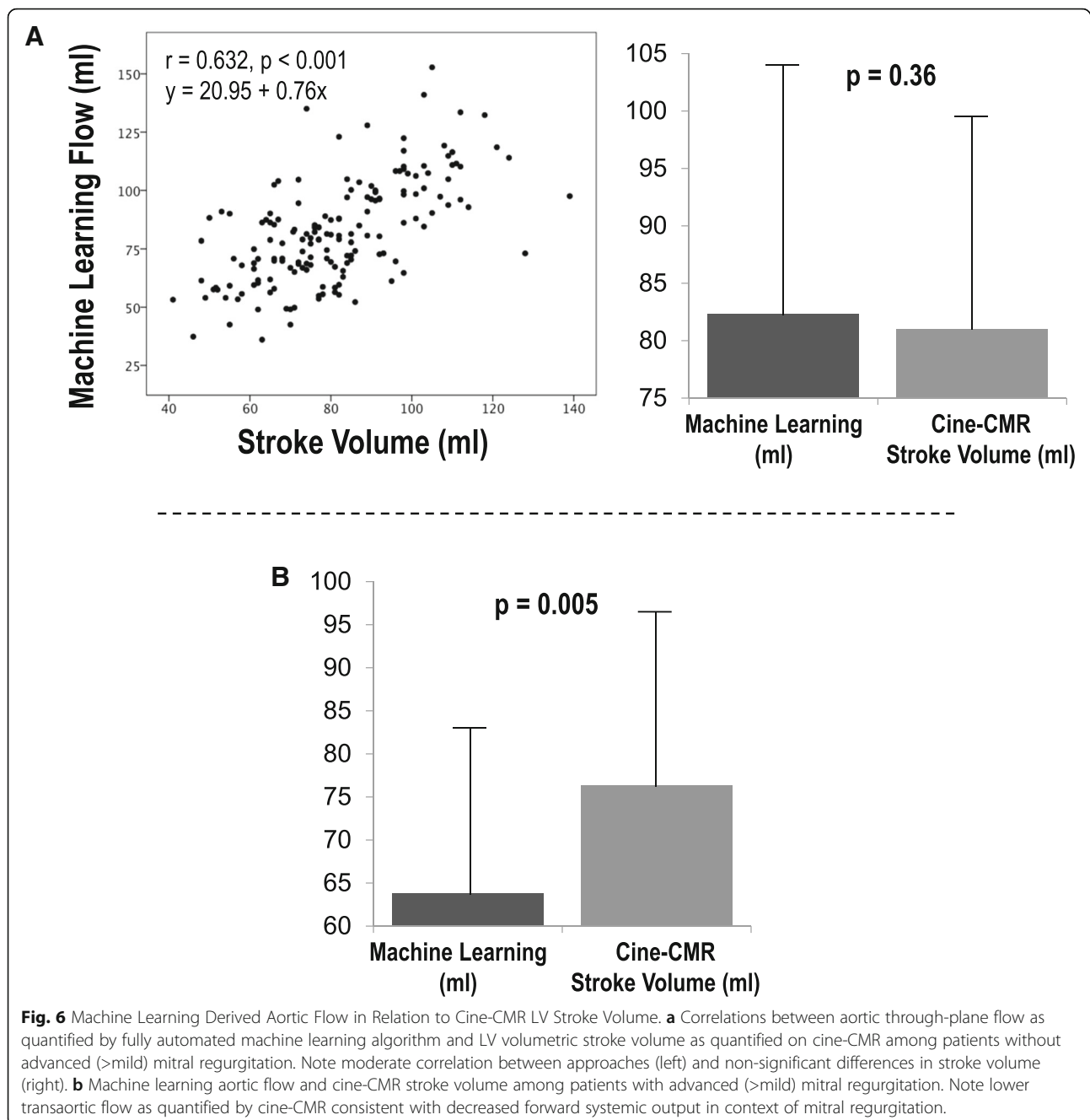
^aBoth $p < 0.01$ (segmentation method vs. manual)^b $p < 0.01$ (machine learning vs. conventional segmentation in terms of MAD)

cumulative time 1.2 min [600-fold shorter than manual segmentation]), as well as close correlations ($r = 0.99$) and small mean differences (1.85 ± 1.80 mL) in flow as quantified manually by an expert CMR reader. Second, machine learning outperformed a conventional automated segmentation algorithm, yielding lower error and error variance with respect to manual segmentation (1.85 ± 1.80 vs 3.33 ± 3.18 mL). Third, machine learning was robust with respect to data source; model performance was excellent when applied to two independent institutional datasets (inclusive of CAD and aortic valve pathology patients) acquired on different CMR scanners. Fourth, machine learning aortic PC-CMR segmentation yielded good correlation ($r = 0.63, p < 0.001$) and non-significant mean differences with cine-CMR derived LV stroke volume in patients without advanced MR: Among patients with advanced mitral regurgitation, machine learning-derived flow was lower than was volumetric LV stroke volume ($\Delta 12.6 \pm 20.9$ mL, $p = 0.005$), further supporting physiologic validity of aortic flow as quantified by our automated segmentation algorithm.

It is important to note that the machine learning approach developed in our study requires no human supervision apart from quality control. In this regard, these results build upon prior work by our group, which has developed fully automated segmentation algorithms for cine-CMR that have been shown to yield superior performance to manual and conventional automated segmentation with respect to clinical robustness, as well as agreement with ex-vivo volumetric phantoms and necropsy-evidenced myocardial mass [23, 24]. In this context, we believe our findings are of substantial relevance to CMR application in large scale-population based datasets.

To the best of our knowledge, this is the first study to apply machine learning for flow quantification on PC-CMR. Prior work has applied deep learning segmentation to the brain, [25, 26] prostate, [27, 28] breast, [29] pulmonary, [7] and musculoskeletal systems [30]. Prior applications of deep learning to CMR have focused on cardiac function/remodeling [8–14], for which analysis has entailed segmentation of cine-CMR (SSFP) datasets. Novelty of our study stems from several different factors: First, our machine learning derived algorithm is focused on PC-CMR - a pulse sequence that is widely applied in clinical practice for which no published CNN-based computer vision model has previously been applied. Second, our model was extensively validated in a broad clinical cohort (inclusive of patients with aortic pathology) imaged at two independent medical centers using CMR scanner equipment from two different vendors – providing confidence that our study findings are robust and broadly applicable. Our model architecture is informed by prior approaches to deep learning analysis of CMR, with specific design and hyperparameter choices tailored to our particular task and dataset. The model makes use of a two-dimensional U-net architecture [16], which applies skip connections between the contracting and expanding pathways of the network to recover fine-grained imaging features during decoding. Residual connections [17] were added to improve the performance of the network by use of identity mappings, which prevent vanishing and exploding gradients during training. Network size was chosen to strike a balance between





expressivity and overfitting given the modest size of our dataset.

Major advantages of the proposed segmentation model are speed and reproducibility. PC-CMR interpretation is time consuming because it requires painstaking manual segmentation. Even with the aid of existing conventional automated methods, substantial manual corrections are often necessary, negating much of the gained efficiency. Further, manual segmentation is prone to subjective decision-making, which adds to quantitative variability and reduces diagnostic confidence. These drawbacks limit

feasibility of large-scale prospective CMR studies, as well as batch analysis of de-identified web-based multicenter registry data (currently > 50,000 exams) currently being accrued by several organizations [31]. The machine learning algorithm tested in this study is well-suited for such applications, given that it is highly reproducible, with speed and accuracy sufficient to process large volumes of data with minimal supervision. Another advantage is the versatility of neural networks with respect to training data. The model described here makes no physical or anatomic assumptions; it simply reproduces the patterns observed

in the training set. As a result, with a sufficient volume of data, the model could easily be repurposed to evaluate other populations, valves, and pathologies.

Several limitations should be noted. First is the lack of a gold-standard test with which to evaluate model performance. All metrics were compared with manual segmentation, which is a suboptimal standard given that it is subjective and stochastic. On the other hand, it should be noted that beyond excellent agreement with manual PC-CMR analysis, machine learning derived data was consistent with two independent imaging approaches; volumetric stroke volume as measured on cine-CMR, as well as differential LV volumetric and aortic forward flow among patients with echocardiography evidenced mitral regurgitation.

It is also important to recognize that in current clinical practice, commercial PC-CMR automated segmentation results are reviewed and frequently manually corrected prior to flow calculation. PC-CMR segmentation tools intended for fully automated segmentation are not developed, and thus could not be used as a comparator in our study. We acknowledge that differences between commercial and manual segmentation would likely have been smaller if commercial segmentation analyses were manually corrected on a frame by frame basis. However, this approach would have prohibited direct comparison between standard commercial segmentation technology and our machine learning derived algorithm, which was a goal of our study. More broadly, manual correction following initial automated segmentation is impractical for analysis of large scale CMR datasets providing a strong rationale for development of more accurate, fully automated, solutions as can be fostered by machine learning. Finally, our training dataset was relatively small in size ($n = 150$ [derivation] and $n = 200$ [derivation+external validation]) and included a select group of patients with CAD and/or aortic valve disease. This is due to the fact that manual segmentation is time- and resource-intensive, which makes it difficult to generate data at the scale one expects in large commercial and industrial settings. Ongoing analysis (informed by current findings) is focused on diagnostic and prognostic applications of this machine learning PC-CMR algorithm in broader cohorts.

Conclusions

The current study provides proof of concept concerning utility of a fully automated deep learning method for PC-CMR aortic flow quantification, demonstrating it to be fast, robust, and superior to that of existing commercial software. The deep learning method described in this manuscript increases interpretive efficiency and reproducibility while maintaining high accuracy, enabling large scale population studies to be performed with

minimal supervision. Future work will focus on expanding machine learning to other cardiac valves, and will compare machine learning and manual segmentation to predict clinical prognosis and therapeutic response in patients undergoing CMR.

Additional file

Additional file 1: Volume Overlap and Surface Distance Equations. (DOCX 14 kb)

Abbreviations

AS: Aortic stenosis; ASSD: Average symmetric surface distance; BAV: Bicuspid aortic valve; CAD: Coronary artery disease; CMR: Cardiovascular magnetic resonance; CNN: Convolutional Neural Networks; HD: Hausdorff Distance; ICC: Interclass correlation coefficient; LV: Left ventricle/left ventricular; LVEF: Left ventricular ejection fraction; MAE: Mean Absolute Error; PC: Phase contrast; VENC: Velocity Encoding Factor

Acknowledgements

Not applicable.

Funding

National Institutes of Health, grant #s R01HL128278-01 (JWW, JK), K23HL102249 (JWW); NIH 1 K23 HL140092-01 (JK).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

Author contributions are as follows: conception and study design (AB, JWW), development of algorithms and analysis software (AB), data collection and protocol design (AB, MP, NC, NHT, RPC, JA, WC, JK, CFA, RJK, JWW), data analysis (AB, MP, JWW), interpretation of data and results (AB, MP, JWW), drafting (AB, MP, JWW), revising (MP, JK, IRD, JWW). All authors read and approved the final manuscript.

Ethics approval and consent to participate

Data usage for this study was approved by the Institutional Review Boards at New York Presbyterian Hospital/Weill Cornell Medicine and Duke University. Subjects in the Cornell validation cohort provided written informed consent for research participation; subjects in the Duke validation cohort comprised use of pre-existing CMR data for patients undergoing imaging for clinical purposes.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Radiology, Weill Cornell Medicine, 525 E 68th St, New York, NY 10065, USA. ²Greenberg Division of Cardiology, Department of Medicine, Weill Cornell Medicine, 525 E 68th St, New York, NY 10065, USA. ³IBM TJ Watson Research Center, 1101 Kitchawan Rd, Yorktown Heights, NY 10598, USA. ⁴Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10065, USA. ⁵Duke Cardiovascular Magnetic Resonance Center, 10 Duke Medicine Circle, Durham, NC 27710, USA. ⁶Weill Cornell Medical College, 525 East 68th Street, New York, NY 10021, USA.

Received: 29 May 2018 Accepted: 18 November 2018

Published online: 07 January 2019

References

- Caruthers SD, Lin SJ, Brown P, Watkins MP, Williams TA, Lehr KA, et al. Practical value of cardiac magnetic resonance imaging for clinical quantification of aortic valve stenosis: comparison with echocardiography. *Circulation*. 2003;108:2236–43.
- Kutty S, Whitehead KK, Natarajan S, Harris MA, Wernovsky G, Fogel MA. Qualitative echocardiographic assessment of aortic valve regurgitation with quantitative cardiac magnetic resonance: a comparative study. *Pediatr Cardiol*. 2009;30:971–7.
- Hundley WG, Li HF, Lange RA, Pfeifer DP, Meshack BM, Willard JE, et al. Assessment of left-to-right intracardiac shunting by velocity-encoded, phase-difference magnetic resonance imaging. A comparison with oximetric and indicator dilution techniques. *Circulation*. 1995;91:2955–60.
- Beerbaum P, Körperich H, Barth P, Esdorn H, Gieseke J, Meyer H. Noninvasive quantification of left-to-right shunt in pediatric patients: phase-contrast cine magnetic resonance imaging compared with invasive oximetry. *Circulation*. 2001;103:2476–82.
- Uretsky S, Gillam L, Lang R, Chaudhry FA, Argulian E, Supariwala A, et al. Discordance between echocardiography and MRI in the assessment of mitral regurgitation severity. a prospective multicenter trial *J Am Coll Cardiol*. 2015;65:1078–88.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–10.
- Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med*. 2018;0. [https://doi.org/10.1016/S2213-2600\(18\)30286-8](https://doi.org/10.1016/S2213-2600(18)30286-8).
- Vigneault DM, Xie W, Ho CY, Bluemke DA, Noble JA. Ω -net (omega-net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med Image Anal*. 2018;48:95–106.
- Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Human-level CMR image analysis with deep fully convolutional networks. arXiv: 171009289 [cs]. 2017. <http://arxiv.org/abs/1710.09289>. Accessed 14 Apr 2018.
- Lieman-Sifry J, Le M, Lau F, Sall S, FastVentricle GD. Cardiac segmentation with ENet. In: Pop M, Wright GA, editors. *Functional imaging and modelling of the heart*: Springer International Publishing; 2017. p. 127–38.
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
- Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng P, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. 2018;1–1.
- Jang Y, Hong Y, Ha S, Kim S, Chang H-J. Automatic segmentation of LV and RV in cardiac MRI. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, et al, editors. *Statistical atlases and computational models of the heart*. ACDC and MMWHS challenges: Springer International Publishing; 2018. p. 161–9.
- Baumgartner CF, Koch LM, Pollefeys M, Konukoglu E. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, et al, editors. *Statistical atlases and computational models of the heart*. ACDC and MMWHS challenges: Springer International Publishing; 2018. p. 111–9.
- Han X. Automatic liver lesion segmentation using a deep convolutional neural network method. 2017. <http://arxiv.org/abs/1704.07239>. Accessed 2 Sep 2018.
- Ronneberger O, Fischer P, U-Net BT. Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention – MICCAI 2015*: Springer International Publishing; 2015. p. 234–41.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv: 1512.03385 [cs]. 2015. <http://arxiv.org/abs/1512.03385>. Accessed 13 Jan 2018.
- 3D Slicer. <https://www.slicer.org/>. Accessed 13 Jan 2018.
- Franco AD, Kim J, Rodriguez-Diego S, Khalique O, Siden JY, Goldberg SR, et al. Multiplanar strain quantification for assessment of right ventricular dysfunction and non-ischemic fibrosis among patients with ischemic mitral regurgitation. *PLoS One*. 2017;12:e0185657.
- Weinsaft JW, Kim J, Medicherla CB, Ma CL, Codella NCF, Kukar N, et al. Echocardiographic algorithm for post-myocardial infarction LV Thrombus. A Gatekeeper for Thrombus Evaluation by Delayed Enhancement CMR *JACC Cardiovasc Imaging*. 2016;9:505–15.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10.
- Jones E, Oliphant T, Peterson P, others. SciPy: Open source scientific tools for Python. 2001. <http://www.scipy.org/>.
- Codella NCF, Cham MD, Wong R, Chu C, Min JK, Prince MR, et al. Rapid and accurate left ventricular chamber quantification using a novel CMR segmentation algorithm: a clinical validation study. *J Magn Reson Imaging*. 2010;31:845–53.
- Codella NCF, Lee HY, Fieno DS, Chen DW, Hurtado-Rua S, Kochar M, et al. Improved left ventricular mass quantification with partial voxel interpolation – in-vivo and necropsy validation of a novel cardiac MRI segmentation algorithm. *Circ Cardiovasc Imaging*. 2012;5:137–46.
- Chen H, Dou Q, Yu L, Qin J, Heng P-A. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*. 2018; 170:446–55.
- Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61–78.
- Cheng R, Roth HR, Lay N, Lu L, Turkbey B, Gandler W, et al. Automatic MR prostate segmentation by deep learning with holistically-nested networks. *International Society for Optics and Photonics*. 2017:101332H. <https://doi.org/10.1117/12.2254558>.
- Tian Z, Liu L, Zhang Z, Fei B. PSNet: prostate segmentation on MRI based on a convolutional neural network. *JMI, JMIOBU*. 2018;5:021208.
- Dalmış MU, Litjens G, Holland K, Setio A, Mann R, Karssemeijer N, et al. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med Phys*. 2017;44:533–46.
- Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a Triplanar convolutional neural network. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 246–53. https://doi.org/10.1007/978-3-642-40763-5_31.
- SCMR Global Registry - Society for Cardiovascular Magnetic Resonance <http://scmr.org/?page=GlobalRegistry>. Accessed 7 May 2018.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

