

RESEARCH

Open Access



Training and clinical testing of artificial intelligence derived right atrial cardiovascular magnetic resonance measurements

Faisal Alandejani¹, Samer Alabed^{1,2}, Pankaj Garg³, Ze Ming Goh¹, Kavita Karunasaagarar⁴, Michael Sharkey^{1,4}, Mahan Salehi¹, Ziad Aldabbagh¹, Krit Dwivedi¹, Michail Mamalakis¹, Pete Metherall⁴, Johanna Uthoff⁵, Chris Johns¹, Alexander Rothman^{1,2,6}, Robin Condliffe⁶, Abdul Hameed^{1,6}, Athanasios Charalampopoulos⁶, Haiping Lu^{2,5}, Sven Plein⁷, John P. Greenwood⁷, Allan Lawrie¹, Jim M. Wild^{1,2}, Patrick J. H. de Koning⁸, David G. Kiely^{1,2,6}, Rob Van Der Geest^{8†} and Andrew J. Swift^{1,2*†} 

Abstract

Background: Right atrial (RA) area predicts mortality in patients with pulmonary hypertension, and is recommended by the European Society of Cardiology/European Respiratory Society pulmonary hypertension guidelines. The advent of deep learning may allow more reliable measurement of RA areas to improve clinical assessments. The aim of this study was to automate cardiovascular magnetic resonance (CMR) RA area measurements and evaluate the clinical utility by assessing repeatability, correlation with invasive haemodynamics and prognostic value.

Methods: A deep learning RA area CMR contouring model was trained in a multicentre cohort of 365 patients with pulmonary hypertension, left ventricular pathology and healthy subjects. Inter-study repeatability (intraclass correlation coefficient (ICC)) and agreement of contours (DICE similarity coefficient (DSC)) were assessed in a prospective cohort (n = 36). Clinical testing and mortality prediction was performed in n = 400 patients that were not used in the training nor prospective cohort, and the correlation of automatic and manual RA measurements with invasive haemodynamics assessed in n = 212/400. Radiologist quality control (QC) was performed in the ASPIRE registry, n = 3795 patients. The primary QC observer evaluated all the segmentations and recorded them as satisfactory, suboptimal or failure. A second QC observer analysed a random subcohort to assess QC agreement (n = 1018).

Results: All deep learning RA measurements showed higher interstudy repeatability (ICC 0.91 to 0.95) compared to manual RA measurements (1st observer ICC 0.82 to 0.88, 2nd observer ICC 0.88 to 0.91). DSC showed high agreement comparing automatic artificial intelligence and manual CMR readers. Maximal RA area mean and standard deviation (SD) DSC metric for observer 1 vs observer 2, automatic measurements vs observer 1 and automatic measurements vs observer 2 is $92.4 \pm 3.5 \text{ cm}^2$, $91.2 \pm 4.5 \text{ cm}^2$ and $93.2 \pm 3.2 \text{ cm}^2$, respectively. Minimal RA area mean and SD DSC metric for observer 1 vs observer 2, automatic measurements vs observer 1 and automatic measurements vs observer 2 was $89.8 \pm 3.9 \text{ cm}^2$, $87.0 \pm 5.8 \text{ cm}^2$ and $91.8 \pm 4.8 \text{ cm}^2$. Automatic RA area measurements all showed moderate correlation

*Correspondence: a.j.swift@sheffield.ac.uk

†Rob Van Der Geest and Andrew J. Swift contributed equally to the manuscript

¹ Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

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



with invasive parameters ($r = 0.45$ to 0.66), manual ($r = 0.36$ to 0.57). Maximal RA area could accurately predict elevated mean RA pressure low and high-risk thresholds (area under the receiver operating characteristic curve artificial intelligence = $0.82/0.87$ vs manual = $0.78/0.83$), and predicted mortality similar to manual measurements, both $p < 0.01$. In the QC evaluation, artificial intelligence segmentations were suboptimal at 108/3795 and a low failure rate of 16/3795. In a subcohort ($n = 1018$), agreement by two QC observers was excellent, kappa 0.84.

Conclusion: Automatic artificial intelligence CMR derived RA size and function are accurate, have excellent repeatability, moderate associations with invasive haemodynamics and predict mortality.

Keywords: Right atrial area, Cardiovascular magnetic resonance, Convolutional neural networks, Artificial intelligence, Deep learning training, Clinical testing, Repeatability assessment, Mortality prediction

Introduction

Changes in the right atrium (RA) are important to recognise in the evaluation of patients with right ventricular (RV) failure [1–5]. Right atrial pressure (RAP) measured at right heart catheterisation is fundamental to the haemodynamic assessment of RV failure [6, 7] and predicts mortality in patients with pulmonary artery hypertension (PAH) [8, 9].

Accurate and repeatable measurements of cardiac chamber size and function are important for patient management [10]. A number of studies have revealed the prognostic significance of cardiovascular magnetic resonance (CMR) measurements in various cardiopulmonary diseases such as cardiomyopathies, pulmonary arterial hypertension (PAH), heart failure and ischaemic heart disease [11–15]. RA size and function measured by CMR can predict mortality [16–18] and the European Society of Cardiology (ESC) and European Respiratory Society (ERS) guidelines advocate the use of maximal (systolic) RA area for stratification of PAH patients [19].

RA measurements are often made manually on images viewed on patient archive and communication systems (PACS) or dedicated software packages with potential for observer variability. Image analysis tools differ between packages and the analysis does take a small but significant amount of time. With the advent of artificial intelligence (AI), deep learning using convolutional neural networks (CNNs), accurate cardiac chamber segmentations are possible [20–24]. Reference ranges for cardiac structure and function in healthy Caucasian adults from the UK Biobank population cohort were described for all four cardiac chambers using CMR [25]. Automated quality control (QC) in image segmentation was applied to the UK Biobank CMR study via the reverse classification accuracy (RCA) approach to categorize between successful and failed segmentations. This previous work showed that RCA has the potential for accurate and fully automatic segmentation QC on a per-case basis [26]. A deep learning based framework for automated, quality-controlled characterization of cardiac function from cine CMR has been established and reference values

for cardiac function metrics were automatically derived from the UK Biobank cohort [27]. Fully automated CMR derived biventricular evaluation of function and morphology in a real-world setting has achieved good results without any operator interaction [28]. However, in the case of unseen anatomic variations, such as severe cardiac chamber shape changes and dilatation as in PAH, or significant artefact, then deep learning measurements may fail or be suboptimal [29].

Automation of RA area measurements may result in lower variability and assist clinicians to reach fast and robust clinical decisions. However, there are currently no studies that have automated CMR RA area metrics in the setting of PAH in which patients have varying degrees of RV failure, and the repeatability, correlation with invasive haemodynamics and success/failure rate in clinical populations remains unknown.

The aim of this study was to develop a quantitative CMR-based automated artificial intelligence (AI) analysis of the RA in a large cohort of patients with heart failure and PAH with varying aetiology and disease severity, and (i) determine the failure rate of the model in a large clinical registry, (ii) evaluate interstudy repeatability, (iii) directly compare the association of manual RA area and AI RA area with invasive haemodynamics and (iv) evaluate RA measurements as predictors of mortality.

Methods

Study population

A cohort of 365 subjects was used for training. This included a random selection of studies from 285 patients in the ASPIRE registry (several ASPIRE follow up scans were included with a total number of studies of 367). Sixty-six subjects from Leeds, including 29 healthy subjects and 37 patients with myocardial infarction of which 19 were acute and 18 were chronic. Fourteen healthy subjects from Leiden University Medical Centre (LUMC) were also included. The total number of studies included in the training cohort was 447. The demographics of the Leeds and Leiden subjects have been previously described [30, 31].

To test the model we used two populations. The first population included 36 patients CMR studies for prospective repeatability testing from the RESPIRE study (ClinicalTrials.gov Identifier: NCT03841344) [32]. The second population contained 400 patients CMR studies for clinical testing from the ASPIRE registry (ASPIRE, ref: c06/Q2308/8). For quality control and failure rate we included 3795 patients (5756 CMR studies, as follow up studies were included) from the ASPIRE registry (Fig. 1). Prospectively recruited patients provided written informed consent. Consent was waived for analysis of retrospective cases.

CMR protocol

The training cohort included 1.5T (HDx, General Electric Healthcare, Chicago, Illinois, USA) and 1.5T (Ingenia, Philips Healthcare, Best, the Netherlands) studies. The testing cohort consisted of GE studies acquired in a clinical setting in the ASPIRE registry. The RESPIRE prospective cohort consisted of GE studies [32]. CMR studies in the testing cohort were performed using a whole-body scanner at 1.5T (HDx (General Electric Healthcare) [33]. Cine CMR acquisitions were made using a balanced steady state free precession (bSSFP) sequence. Following planning sequences, 4-chamber cine images were acquired. A stack of short axis images were acquired covering apex to base. Slice thickness and number of cardiac phases were 8 mm with 20 phases.

Leeds and Leiden CMR studies were performed on a 1.5 T system (Ingenia, Philips Healthcare) equipped with a 28-channel flexible torso coil and digitization of the CMR signal in the receiver coil. Vertical long-axis, horizontal long-axis, 3-chamber (left ventricular (LV) outflow tract-views), and the LV volume contiguous short axis stack cine imaging were defined using survey. All cines

were acquired with a bSSFP, single-slice breath-hold sequence. Typical parameters for bSSFP cine were as follows: SENSE factor 2, flip angle 60°, TE 1.5 ms, TR 3 ms, field of view 320–420 mm according to patient size, slice thickness 8 mm and 30 phases per cardiac cycle.

Image analysis

Four observers SA, FA, KK and AJS (with 2, 3, 13 and 11 years CMR experience, respectively) manually drew LV and RV and atrial contours in 4-chamber cine CMR views on all cardiac phases for the training and testing cohorts. All contours were drawn with observers blinded to the patient’s clinical information. All manual contours were reviewed by an expert CMR reader (AJS). RV endocardial and epicardial surfaces were also manually traced from the stack of short-axis cine images to obtain RV volumetric and functional measurements as previously described [33]. MASS software (research version 2020; Leiden University Medical Center, Leiden, the Netherlands) was used for the manual contouring for developing the algorithm and repeatability testing).

Deep learning training

CMR studies including a random selection of patients from the ASPIRE registry, subjects from Leeds, and from LUMC were used for deep learning training. The training process was performed in two stages. We trained two CNN models with different numbers of manually annotated 4-chamber view images in the training set. The validation set and test set used were the same for both of the CNN models. Since no hyper parameter tuning was performed in the current experiments a relatively small validation set of 6 subjects (180 images) was deemed sufficient to confirm model convergence during training and to confirm that the models did not suffer from

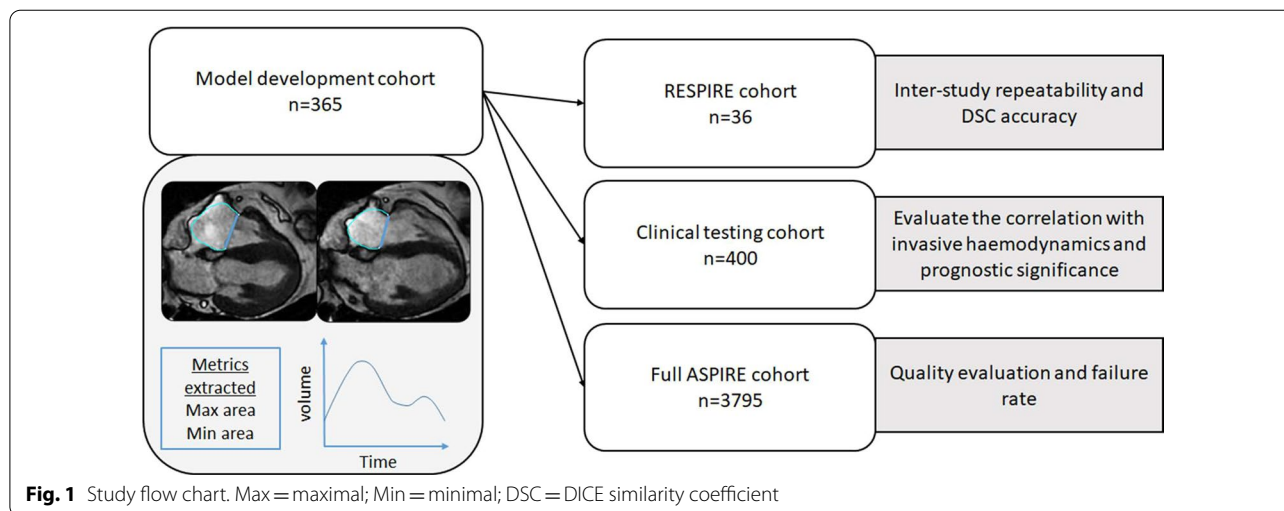


Fig. 1 Study flow chart. Max = maximal; Min = minimal; DSC = DICE similarity coefficient

overfitting. The test set consisting of 20 cases was used to compare the model performance of the initial model with the final model. Following this strategy we maximised the number of studies available for training. The initial model was trained on a combination of Philips (Leeds/LUMC, $n=80$) and GE (Sheffield, $n=184$) data (total $n=264$). The contours used for training were all generated without the use of a CNN. For the final model 183 additional Sheffield GE scans were added. The contours for these additional cases were generated by reviewing and editing the contours generated using the base model. On average 50% of the contours generated by the initial model were manually edited for this set of cases. These cases were separate from the test cohorts.

The CNNs used for the experiments had an UNET-like architecture with 16 convolutional layers including residual learning units and was implemented using Python and TensorFlow. Input images were resampled to a fixed pixel spacing of 1 mm and cropped to a 256×256 image matrix size and zero filled when required. During training, data augmentation was performed on the fly by creating new training samples by randomly rotating, flipping, shifting and modifying image intensities of the original images. A total of 447 manually annotated 4-chamber cine series were used for training corresponding to 10,045 images. For training the Adam optimizer method was used, the learning rate was selected as 0.001 and cross-entropy was used as loss function. Each training batch included a random selection of 20 images. The number of epochs was set at a fixed number of 50, with all images used once in every epoch. The raw output of the CNNs is a labeled image, with the six possible label values corresponding to either one of the four cardiac cavities, the LV myocardium, or background. For each cardiac label, the largest connected component was extracted and a closed spatially smoothed contour around the extracted region generated. The area of the cardiac cavities was subsequently derived as the area surrounded by the generated contours. All experiments were executed on a standard PC with Intel Core i7 CPU with 64 GB of internal RAM memory equipped with an Nvidia GTX 1080 TI GPU with 12 GB of memory. The authors are happy to be contacted for research access to the Mass software and the AI segmentation tool upon request.

Quality control

All automatically AI segmented RA area contours across all cardiac phases and resultant volume-time curves were evaluated by AS and scored as satisfactory, suboptimal or failure. In addition, the quality of the image acquisition was assessed for artefacts and slice position error. The definitions for QC were assigned prior to image review. Satisfactory was defined as either perfect contouring or

minor errors that were not thought to affect the volumetric results. Suboptimal was defined as contours with errors deemed significant enough to affect the volumetric results. Failure defined as either absent contours or gross failure of the algorithm to segment the cardiac structures.

Repeatability and agreement of the deep learning contours

To evaluate inter-study agreement two CMR scans were performed on the same day in two separate sittings as part of the RESPIRE study [32] for AI and manual measurements. In addition, interobserver agreement assessments, manual (AS) vs manual (FA), AI vs AS and AI versus FA were made. Agreement of the machine learning contouring model was evaluated by DSC. The DICE similarity for all cardiac cavities was computed in the 20 subjects in the test set. This was both for the baseline model as well as the final model.

Association of manual and AI CMR measurements with invasive haemodynamics

Correlations with invasive haemodynamics were performed in patients in the ASPIRE registry clinical testing cohort who underwent right heart catheterisation within 48 h of CMR. The accuracy of RA CMR measurements to predict ESC/ERS mean RAP low and high-risk thresholds of 8 mmHg and 14 mmHg respectively, was assessed.

Statistical analysis

Continuous variables are presented as proportions and means \pm standard deviations. Normal distribution assessed by visual inspection of histograms and using the Shapiro–Wilk test. Variables that were not normally distributed were correlated using Spearman correlation coefficient. Univariate Cox regression Hazard ratios were calculated for AI and manual RA measurements to estimate the prognostic significance. Accuracy of RA measurements to predict RA thresholds performed using receiver operating characteristic analysis. Intraclass correlation coefficients and Bland–Altman plots were used to assess repeatability of manual and AI CMR metrics. Inter-rater reliability of the two observers grading of segmentation quality as satisfactory, suboptimal or failure was assessed using Cohen's kappa testing in a subcohort. Statistical analysis was carried out using SPSS (version 26, Statistical Package for the Social Sciences, International Business Machines, Inc., Armonk, New York, USA) and RStudio (version 1.2.5033, RStudio, Boston, Massachusetts, USA), and p value of 0.05 was considered statistically significant. For data presentation, GraphPad Prism (version 9.1.0, GraphPad Software, San Diego, California, USA) software was used.

Results

Patients

The ASPIRE registry in the training model included patients with left heart disease (15%), lung disease (12%), chronic thromboembolic PAH (21%), PAH (29%), other PAH (2%) and non-PAH (21%). The mean and standard deviation (SD) of the main haemodynamics of the ASPIRE registry in the training model is 10.4 ± 6.2 mmHg for mean RAP, 41.0 ± 15.5 mmHg for mean pulmonary arterial pressure, 13.4 ± 6.0 mmHg for pulmonary arterial wedge pressure, and 561 ± 466 dynes/

m2 for pulmonary vascular resistance. The characteristics for the prospective repeatability, clinical testing and full cohort are presented in Table 1. In the clinical testing cohort, 218 of the 400 patients had died (54.5%) during a mean follow-up period of 1 year.

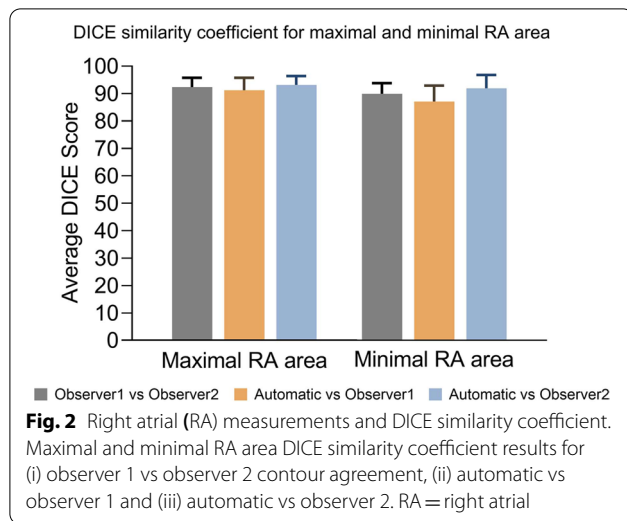
Quality control

Of 3795 patients (5756 studies) analysed by the AI model, 16 (0.3%) failed. 108 (1.9%) had suboptimal contours significant enough to be thought to affect the area measurements. In 72/108 patients, the 4-chamber slice was

Table 1 Demographics, CMR and invasive haemodynamics of patients in the (i) RESPIRE (ii) Clinical testing and (iii) full cohort

	RESPIRE repeatability (n = 36)	Clinical testing (n = 400)	Full cohort (n = 3795)
Demographics			
Age, yr	49.5 ± 15.9	55.4 ± 16.4	62.8 ± 15.3
Sex, F/M (F %)	30/6 (83)	283/117 (71)	2355/1440 (62)
BSA (m ²)	1.9 ± 0.2	1.8 ± 0.2	1.8 ± 0.2
WHO FC I, n (%)	0 (0)	2 (1)	47 (1)
WHO FC II, n (%)	2 (6)	21 (5)	441 (12)
WHO FC III, n (%)	30 (83)	338 (85)	2743 (77)
WHO FC IV, n (%)	4 (11)	36 (9)	336 (10)
Diagnosis, n (%)			
Left Heart Disease	0 (0)	0 (0)	611 (16)
Lung Disease	0 (0)	0 (0)	632 (17)
CTEPH	0 (0)	0 (0)	728 (19)
PAH	36 (100)	400 (100)	1040 (28)
Other PAH	0 (0)	0 (0)	84 (2)
Other (not PAH)	0 (0)	0 (0)	677 (18)
Haemodynamics			
mRAP, mmHg	11 ± 7	10.4 ± 6.0	10.1 ± 6.0
mPAP, mmHg	52 ± 13	48.0 ± 13.7	40.8 ± 14.2
PAWP, mmHg	10 ± 3	10.3 ± 2.9	12.8 ± 5.9
Cardiac output L/min	4.5 ± 1.7	4.9 ± 1.5	4.9 ± 1.9
Cardiac index, L/min/m ²	2.5 ± 0.9	2.8 ± 0.9	2.7 ± 1.0
PVR, dynes/m ²	899 ± 512	720 ± 419	562 ± 419
MvO ₂ , %	65.0 ± 9.1	63.5 ± 9.1	65.2 ± 9.3
CMR volumetric measurements			
RVESVI, ml/m ²	25.4 ± 9.2	46.8 ± 28.2	37.3 ± 27.1
RVEDVI, ml/m ²	63.3 ± 27.6	72.7 ± 35.5	62.6 ± 35.5
RVSVI, ml/m ²	37.9 ± 20.7	25.9 ± 12.7	25.3 ± 15.4
RVEF, %	43.3 ± 10.0	39.1 ± 14.1	44.6 ± 16.1
CMR area measurements			
Automatic max RA area, cm ²	22.6 ± 6.3	25.5 ± 9.8	25.8 ± 10.6
Manual max RA area, cm ²	22.5 ± 6.3	26.0 ± 10.3	-
Automatic min RA area, cm ²	15.0 ± 5.5	18.4 ± 9.4	18.5 ± 10.3
Manual min RA area, cm ²	15.3 ± 5.7	19.3 ± 10.1	-

BSA, body surface area; CMR, cardiovascular magnetic resonance; CTEPH, chronic thromboembolic pulmonary hypertension; max, maximal; min, minimal; mRAP, mean right atrial pressure; mPAP, mean pulmonary arterial pressure; MvO₂, mixed venous oxygen saturation; PAH, pulmonary arterial hypertension; PAWP, pulmonary arterial wedge pressure; PH, pulmonary hypertension; PVR, pulmonary vascular resistance; RHC, right heart catheterization; RVESVI, right ventricular end-systolic volume index; RVEDVI, right ventricular end-diastolic volume index; RVSVI, right ventricular stroke volume index; RVEF, right ventricular ejection fraction; RA, right atrial; WHO FC, World Health Organisation functional class. Data presented as mean ± standard deviation



off-plane, with the most frequent error being inclusion of the LV outflow tract and suboptimal view of the RA. In 36/108 severe image artefact, typically breathing artefact or poor cardiac gating lead to suboptimal RA contours. In a randomly selected subcohort of 1018 studies, the scoring of satisfactory, suboptimal and failure showed excellent agreement between observer 1 and observer 2, with a high kappa statistic of 0.84.

Segmentation agreement

Manual and automatic AI segmentation were assessed in the same day repeat studies from the prospective RESPIRE study. DSC showed high agreement (Fig. 2) comparing automatic AI and manual CMR readers, with

a minimal bias towards either reader, validating similarity in the resulting contours. Manual contours made by observer 1 and observer 2 were closely related for both maximal RA area and minimal RA area. The mean and SD DSC metric for observer 1 vs observer 2, AI measurements vs observer 1 and AI measurements vs observer 2 is 92.4 ± 3.5 , 91.2 ± 4.5 and 93.2 ± 3.2 for maximal RA area. The mean and SD DSC metric for observer 1 vs observer 2, AI measurements vs observer 1 and AI measurements vs Observer 2 is 89.8 ± 3.9 , 87.0 ± 5.8 and 91.8 ± 4.8 for minimal RA area. The DSC for all four cardiac chambers before and after refinement for the 20 subjects in the test set are shown in Additional file 1: Table S1.

Repeatability and agreement assessment

All AI RA measurements showed higher interstudy (scan-rescan) repeatability ICC 0.91 to 0.95, compared to manual measurements (observer 1 ICC 0.82 to 0.88, observer 2 ICC 0.88 to 0.91). Similar repeatability was also found comparing both observers with AI RA contours compared to observer 1 vs observer 2 ICC 0.96 to 0.98, see Tables 2, 3. Minimal bias was found for AI RA measurements, Fig. 3.

Clinical testing cohort

In the clinical testing cohort (n = 400), RA area measurements made by AI and observers were comparable (Table 1). In the clinical testing cohort both manual and AI maximal RA area predicted overall all-cause mortality with similar predictive value, (hazard ratio 1.02 (95% confidence interval 1.01 to 1.03) and 1.02 (95% confidence interval 1.01 to 1.03) respectively, both $p < 0.01$).

Table 2 Scan-rescan variability of automatic AI and manual right atrial CMR measurements

	Automatic		Observer 1		Observer 2	
	ICC	95% CI	ICC	95% CI	ICC	95% CI
	Max RA area	0.91	0.82, 0.96	0.82	0.65, 0.91	0.88
Min RA area	0.95	0.89, 0.97	0.88	0.75, 0.94	0.91	0.84, 0.96

Table 3 Interobserver variability of automatic AI and manual right atrial CMR measurements

	Automatic vs Observer 1		Automatic vs Observer 2		Observer 1 vs Observer 2	
	ICC	95% CI	ICC	95% CI	ICC	95% CI
	Max RA area	0.99	0.97, 0.99	0.98	0.95, 0.99	0.98
Min RA area	0.99	0.98, 0.99	0.97	0.92, 0.99	0.96	0.95, 0.99

AI, artificial intelligence; CMR, cardiovascular magnetic resonance; max, maximal; min, minimal; RA, right atrial

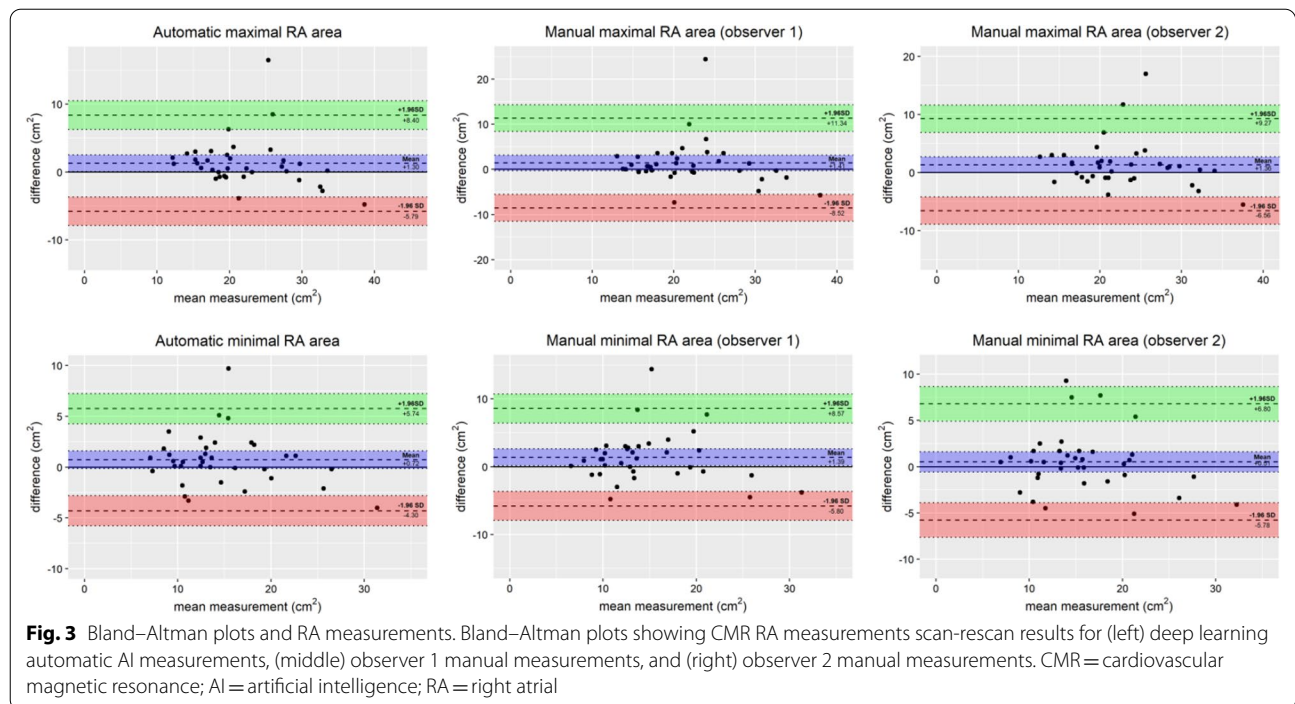


Table 4 Pearson correlation (*r*) for the relation of manual maximal RA area and automatic AI maximal RA area with RHC parameters. mRAP, mean right atrial pressure; PVR, pulmonary vascular resistance

RHC parameters	Manual maximal RA area (n = 212)		Automatic maximal RA area (n = 212)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>P</i>
mRAP	0.57	<0.001	0.64	<0.001
mPAP	0.38	<0.001	0.46	<0.001
Cardiac index	−0.36	<0.001	−0.45	<0.001
PVR	0.36	<0.001	0.47	<0.001
SvO2	−0.41	<0.001	−0.48	<0.001

Table 5 Pearson correlation (*r*) for the relation of manual minimal RA area and automatic AI minimal RA area with RHC parameters

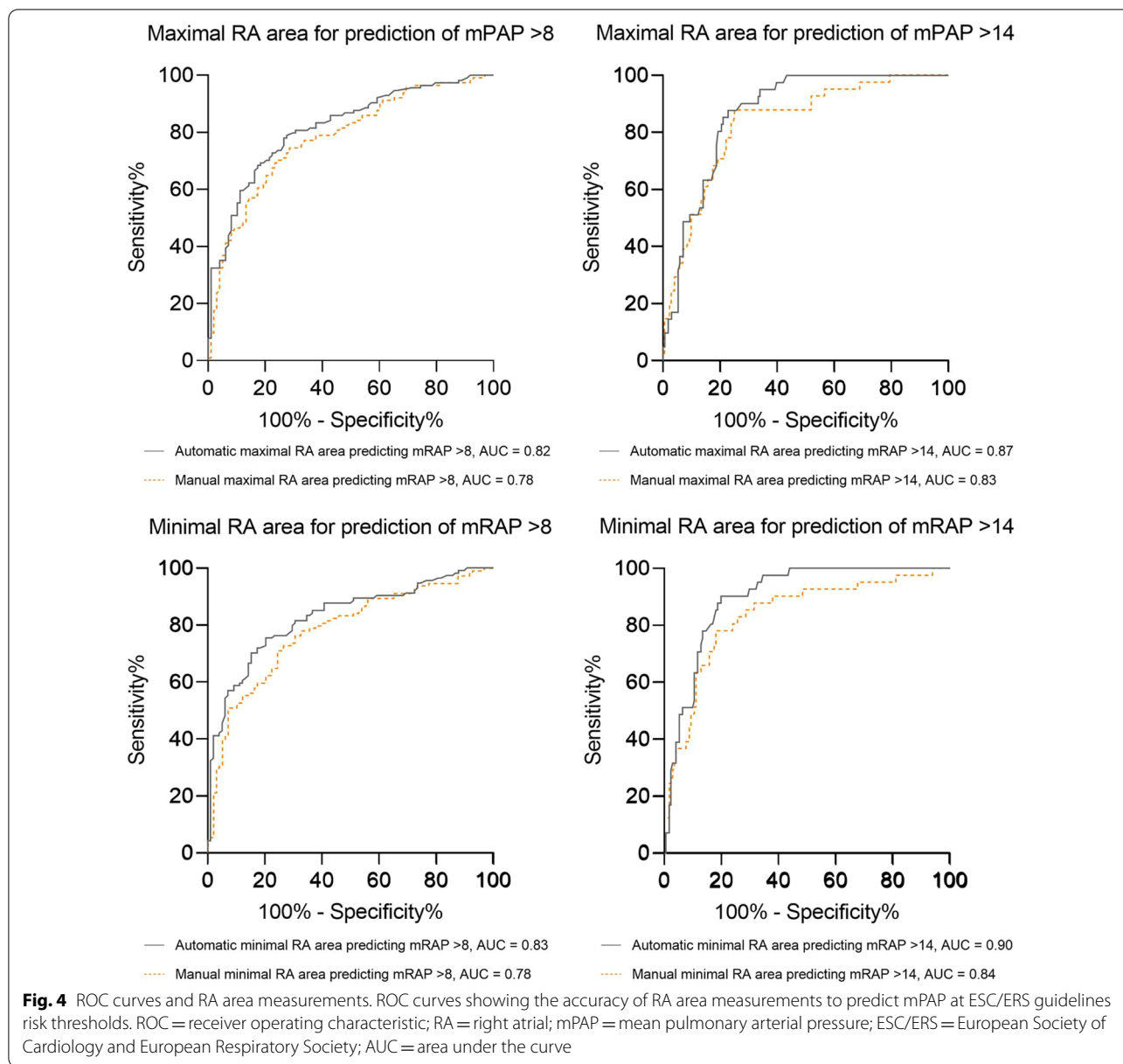
RHC parameters	Manual minimal RA area (n = 212)		Automatic minimal RA area (n = 212)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
mRAP	0.57	<0.001	0.66	<0.001
mPAP	0.40	<0.001	0.50	<0.001
Cardiac index	−0.39	<0.001	−0.50	<0.001
PVR	0.40	<0.001	0.54	<0.001
SvO2	−0.44	<0.001	−0.55	<0.001

RA, right atrial; AI, artificial intelligence; RHC, right heart catheterization; mRAP, mean right atrial pressure; mPAP, mean pulmonary arterial pressure; PVR, pulmonary vascular resistance; MvO2, mixed venous oxygen saturation

Manual and AI minimal RA area also showed a similar predicted mortality hazard ratio of 1.03 (95% confidence interval 1.01 to 1.02) and 1.02 (95% confidence interval 1.01 to 1.03), respectively, both *p* < 0.01.

Of the 400 patients identified for the clinical testing cohort, 212 patients underwent CMR and right heart catheterization (RHC) within 48 h. Moderate positive correlations were found between RA area measurements and mean RAP (mRAP) (AI, *r* = 0.64 and manual, *r* = 0.57). Moderate correlations of AI maximal RA area measurements with all invasive haemodynamics were found, see Table 4. The strongest correlation was found between minimal RA area and mRAP, *r* = 0.66, see Table 5.

Maximal RA area could accurately predict mRAP low and high ESC/ERS risk thresholds (area under the receiver operating characteristic curve AI = 0.82 vs manual = 0.78 to identify low-risk patients with mRAP ≤ 8 mmHg and AI = 0.87 vs manual = 0.83 to identify high-risk patients with mRAP > 14 mmHg). Minimal RA area had a marginally highest accuracy for prediction of elevated mRAP, the strongest prediction was for mPAP > 14, area under the curve (AUC) 0.90, see Fig. 4. In comparison with manual measurements, automatic maximal RA area was not more accurate for detection of patients with mRAP > 8 mmHg and mRAP > 14 mmHg,



($p=0.11$) and ($p=0.13$), respectively. Automatic contouring of minimal RA area trended to suggest higher accuracy for predicting elevated mRAP >8 mmHg and mRAP > 14 mmHg than manual measurements ($p=0.05$) ($p=0.06$), respectively, however these results are not of statistical significance.

Discussion

This study shows that CMR RA area measurements can be fully automated using AI with a very low failure rate in a large clinical cohort with varying RA size and

deformity. The variability of AI derived RA area measurements is lower than manual measurements in a scan-rescan cohort of patients with varying severities of RA size and function, and PAH. RA area measurements moderately correlate with invasive haemodynamics, and AI measurements can identify mRAP prognostic thresholds with more confidence than manual measurements, finally RA area measurements predict mortality with similar accuracy to manual measurements.

This study shows that fully automated AI-based contouring of the RA has a very low AI failure rate of ~2%

in a large clinical population of patients with varying degrees of breathlessness, exercise limitation and aetiology of cardiac and pulmonary disease. The main reasons for failure were severe artefact, in particular poor cardiac gating, image noise and acquisition issues such as poor slice positioning of the 4-chamber slice, the latter the most common scenario. Such images cannot yield accurate RA area measurements by an observer or AI.

Using CMR, reference ranges for cardiac structure and function in healthy adults were previously described for all four cardiac chambers [25]. Automation of the QC process can potentially assist in validating AI algorithms. The potential for accurate and fully automatic segmentation QC has been demonstrated and applied to the UK Biobank CMR study using the RCA approach [26]. Reference values for cardiac function metrics were automatically derived from the UK Biobank and a deep learning based framework for automated, quality-controlled characterization of cardiac function from cine CMR has been confirmed [27]. Although, we advocate use of observer review in the QC process to maintain oversight of the segmented contours.

Assessment of interstudy (scan-rescan) repeatability is crucial to evaluate the utility of imaging measurements [34]. Interstudy repeatability is especially important for the comparison of automatic AI measurements with manual measurements [35]. We utilised a prospective scan-rescan study with rigorous study design [32] and show AI measurements are highly repeatable with marginally higher repeatability than manual measurements. Lower variability has advantages for more precise evaluation of changes in the RA following therapeutic intervention in trials and clinical practice, where treatment decisions are impacted by progressive structural and functional changes in the heart.

The ASPIRE registry includes a wide range of pathology including PAH, left heart failure, lung disease, chronic thromboembolic disease and patients found to have normal invasive haemodynamics. The AI 'seeing' a wider range of pathology is of paramount importance [20]. This is the first study to compare AI and manual measurements with invasive haemodynamic measurements of RAP. Here in this diverse population we identify a close correlation of AI RA area measurements with invasive mRAP, this combined with the low scan-rescan variability supports its potential use as a clinical tool. We show that RA area measurements are prognostic to a similar level as manual measurements. Further work to evaluate AI metrics in risk stratification is required as has been achieved for RV measurements [33]. In addition further work will be to clinically evaluate the range of physiological parameters that can be extracted from the

AI segmentations, such as RA strain [36, 37] and potentially reservoir and conduit function [38, 39]. RHC measurements correlated strongly with AI RA measurements, indicating AI metrics may provide physiologically accurate measure of pathophysiological changes in the heart given their high consistency and repeatability.

Limitations and future work

This is a single centre clinical testing of an AI algorithm developed in a multi-vendor multicentre cohort, with the clinical testing in the setting of a tertiary referral centre for patients with PAH. The imaging appearances and patient populations are likely representative of other PAH referral centres. The algorithm was generated in a multicentre setting, with single centre testing. Multicentre testing would be the next step to determine wider applicability of the algorithm. The current approach uses manual QC which is advantageous from a regulatory standpoint and maintains expert oversight of the AI. Future work to automate QC is of interest, however we consider manual review an important component of the system. Furthermore, future work will include evaluation of the utility of such automatic QC approaches in clinical populations.

This study developed an AI model for RA area estimation rather than volume. The rationale was to automate measurements made clinically and consistent with the ESC/ERS guidelines in PAH. Further work to develop and clinically evaluate a 3-dimensional or multislice RA volumetric model would be of value and work to extract physiological parameters previously suggested to be important [17] may be of benefit in future studies. Future work will be to explore the development of a four chamber AI prognostic model in PAH.

Conclusion

In this study we have developed, tested and clinically validated an AI model to fully automate CMR RA area measurements. The data suggests great clinical applicability of AI derived RA measurements, in addition to time saving benefits.

Abbreviations

AI: Artificial intelligence; AUC: Area under the curve; bSSFP: Balanced steady state free precession; CMR: Cardiovascular magnetic resonance; CNNs: Convolutional neural networks; DSC: DICE similarity coefficient; ESC/ERS: European Society of Cardiology and European Respiratory Society; ICC: Intraclass correlation coefficient; LUMC: Leiden University Medical Centre; LV: Left ventricle/left ventricular; mRAP: Mean right atrial pressure; PACS: Patient archive and communication systems; PAH: Pulmonary arterial hypertension; QC: Quality control; RA: Right atrium/right atrial; RAP: Right atrial pressure; RCA: Reverse classification accuracy; RHC: Right heart catheterization; RV: Right ventricle/right ventricular; SD: Standard deviation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12968-022-00855-3>.

Additional file 1: Table S1. DSC values before and after refinement for all four cardiac chambers area.

Acknowledgements

Not applicable.

Authors' contributions

All authors read and approved the final manuscript.

Funding

Andrew Swift is supported by a Wellcome Trust fellowship grant 205188/Z/16/Z. This work was supported by an NIHR AI Award, AI_AWARD01706.

Availability of data and materials

These can be provided upon request to the corresponding author.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained for the prospective repeatability study, [32] (ClinicalTrials.gov Identifier: NCT03841344) and the ASPIRE registry (ASPIRE, ref: c06/Q2308/8). Prospectively recruited patients provided written informed consent. Consent was waived for analysis of retrospective cases.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK. ²INSIGNEO, Institute for In Silico Medicine, University of Sheffield, Sheffield, UK. ³Norwich Medical School, University of East Anglia, Norwich, UK. ⁴Radiology Department, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. ⁵Department of Computer Science, University of Sheffield, Sheffield, UK. ⁶Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. ⁷Multidisciplinary Cardiovascular Research Centre (MCRC) & Biomedical Imaging Science Department, Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Clarendon Way, Leeds, UK. ⁸Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, Netherlands.

Received: 15 September 2021 Accepted: 19 March 2022

Published online: 07 April 2022

References

- Austin C, Allassas K, Burger C, Safford R, Pagan R, Duello K, et al. Echocardiographic assessment of estimated right atrial pressure and size predicts mortality in pulmonary arterial hypertension. *Chest*. 2015;147(1):198–208.
- Raymond RJ, Hinderliter AL, Willis PW, Ralph D, Caldwell EJ, Williams W, et al. Echocardiographic predictors of adverse outcomes in primary pulmonary hypertension. *J Am Coll Cardiol*. 2002;39(7):1214–9.
- Roca GQ, Campbell P, Claggett B, Solomon SD, Shah AM. Right atrial function in pulmonary arterial hypertension. *Circ Cardiovasc Imag*. 2015. <https://doi.org/10.1161/CIRCIMAGING.115.003521>.
- Fukuda Y, Tanaka H, Motoji Y, Ryo K, Sawa T, Imanishi J, et al. Utility of combining assessment of right ventricular function and right atrial remodeling as a prognostic factor for patients with pulmonary hypertension. *Int J Cardiovasc Imaging*. 2014;30(7):1269–77.
- Fukuda Y, Tanaka H, Ryo-Koriyama K, Motoji Y, Sano H, Shimoura H, et al. Comprehensive functional assessment of right-sided heart using speckle tracking strain for patients with pulmonary hypertension. *Echocardiography*. 2016;33(7):1001–8.
- Damman K, van Deursen VM, Navis G, Voors AA, van Veldhuisen DJ, Hillege HL. Increased central venous pressure is associated with impaired renal function and mortality in a broad spectrum of patients with cardiovascular disease. *J Am Coll Cardiol*. 2009;53(7):582–8.
- Drazner MH, Rame JE, Stevenson LW, Dries DL. Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure. *N Engl J Med*. 2001;345(8):574–81.
- Lichtblau M, Bader PR, Saxer S, Berlier C, Schwarz EI, Hasler ED, et al. Right atrial pressure during exercise predicts survival in patients with pulmonary hypertension. *J Am Heart Assoc*. 2020. <https://doi.org/10.1161/JAHA.120.018123>.
- Daloz GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Survival in patients with primary pulmonary hypertension: results from a national prospective registry. *Ann Intern Med*. 1991;115(5):343–9.
- Kiely DG, Levin DL, Hassoun PM, Ivy D, Jone P-N, Bwika J, et al. Statement on imaging and pulmonary hypertension from the Pulmonary Vascular Research Institute (PVRI). *Pulm Circ*. 2019;9(3):1.
- Klem I, Shah DJ, White RD, Pennell DJ, van Rossum AC, Regenfus M, et al. Prognostic value of routine cardiac magnetic resonance assessment of left ventricular ejection fraction and myocardial damage an international. Multicenter Study *Circ Cardiovasc Imag*. 2011;4(6):610–9.
- Mordi I, Bezerra H, Carrick D, Tzemos N. The combined incremental prognostic value of LVEF, late gadolinium enhancement, and global circumferential strain assessed by CMR. *Jacc Cardiovasc Imag*. 2015;8(5):540–9.
- Swift AJ, Capener D, Johns C, Hamilton N, Rothman A, Elliot C, et al. Magnetic resonance imaging in the prognostic evaluation of patients with pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2017;196(2):228–39.
- Rodriguez-Palomares JF, Gavara J, Ferreira-Gonzalez I, Valente F, Rios C, Rodriguez-Garcia J, et al. Prognostic value of initial left ventricular remodeling in patients with reperfused STEMI. *Jacc Cardiovasc Imag*. 2019;12(12):2445–56.
- Alabed S, Shahin Y, Garg P, Alandejani F, Johns CS, Lewis RA, et al. Cardiac-MRI predicts clinical worsening and mortality in pulmonary arterial hypertension: a systematic review and meta-analysis. *JACC Cardiovasc Imag*. 2020;14:931.
- Ivanov A, Mohamed A, Asfour A, Ho J, Khan SA, Chen O, et al. Right atrial volume by cardiovascular magnetic resonance predicts mortality in patients with heart failure with reduced ejection fraction. *PLoS ONE*. 2017;12(4):e0173245.
- Sato T, Tsujino I, Ohira H, Oyama-Manabe N, Ito YM, Yamada A, et al. Right atrial volume and reservoir function are novel independent predictors of clinical worsening in patients with pulmonary hypertension. *J Heart Lung Transplant*. 2015;34(3):414–23.
- Sallach JA, Tang WHW, Borowski AG, Tong W, Porter T, Martin MG, et al. Right atrial volume index in chronic systolic heart failure and prognosis. *Jacc Cardiovasc Imag*. 2009;2(5):527–34.
- Galie N, Humbert M, Vachiery J-L, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS); endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J*. 2015;46(4):903–75.
- Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. 2020;7:25.
- Tao Q, Yan W, Wang Y, Paiman EHM, Shamon DP, Garg P, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine mr images: a multivendor. Multicenter Study *Radiol*. 2019;290(1):81–8.
- Suinesiaputra A, Sanghvi MM, Aung N, Paiva JM, Zemrak F, Funk K, et al. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. *Int J Cardiovasc Imaging*. 2018;34(2):281–91.
- Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*. 2018;20:1.

24. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng P-A, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. 2018;37(11):2514–25.
25. Petersen SE, Aung N, Sanghvi MM, Zemrak F, Fung K, Paiva JM, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J Cardiovasc Magnetic Reson*. 2017. <https://doi.org/10.1186/s12968-017-0327-9>.
26. Robinson R, Valindria VV, Bai W, Oktay O, Kainz B, Suzuki H, et al. Automated quality control in image segmentation: application to the UK Biobank cardiovascular magnetic resonance imaging study. *J Cardiovasc Magnetic Reson*. 2019. <https://doi.org/10.1186/s12968-019-0523-x>.
27. Ruijsink B, Puyol-Anton E, Oksuz I, Sinclair M, Bai W, Schnabel JA, et al. Fully automated, quality-controlled cardiac analysis from CMR validation and large-scale application to characterize cardiac function. *Jacc Cardiovasc Imag*. 2020;13(3):684–95.
28. Backhaus SJ, Staab W, Steinmetz M, Ritter CO, Lotz J, Hasenfuss G, et al. Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: applicability to clinical routine settings. *J Cardiovasc Magnetic Reson*. 2019. <https://doi.org/10.1186/s12968-019-0532-9>.
29. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol*. 2018;15(3):504–8.
30. Garg P, Crandon S, Swoboda PP, Fent GJ, Foley JRJ, Chew PG, et al. Left ventricular blood flow kinetic energy after myocardial infarction—insights from 4D flow cardiovascular magnetic resonance. *J Cardiovasc Magnetic Reson*. 2018. <https://doi.org/10.1186/s12968-018-0483-6>.
31. Crandon S, Westenberg JJM, Swoboda PP, Fent GJ, Foley JRJ, Chew PG, et al. Impact of age and diastolic function on novel, 4D flow CMR biomarkers of left ventricular blood flow kinetic energy. *Sci Rep*. 2018. <https://doi.org/10.1038/s41598-018-32707-5>.
32. Swift AJ, Wilson F, Cogliano M, Kendall L, Alandejani F, Alabed S, et al. Repeatability and sensitivity to change of non-invasive end points in PAH: the RESPIRE study. *Thorax*. 2021;76:1032.
33. Lewis RA, Johns CS, Cogliano M, Capener D, Tubman E, Elliot CA, et al. Identification of cardiac magnetic resonance imaging thresholds for risk stratification in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2020;201(4):458–68.
34. Grothues F, Moon JC, Bellenger NG, Smith GS, Klein HU, Pennell DJ. Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance. *Am Heart J*. 2004;147(2):218–23.
35. Augusto JB, Davies RH, Bhuvana AN, Knott KD, Seraphim A, Alfarihi M, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digital Health*. 2021;3(1):E20–8.
36. Maceira AM, Cosin-Sales J, Prasad SK, Pennell DJ. Characterization of left and right atrial function in healthy volunteers by cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*. 2016. <https://doi.org/10.1186/s12968-016-0284-8>.
37. Xie E, Yu R, Ambale-Venkatesh B, Bakhshi H, Heckbert SR, Soliman EZ, et al. Association of right atrial structure with incident atrial fibrillation: a longitudinal cohort cardiovascular magnetic resonance study from the Multi-Ethnic Study of Atherosclerosis (MESA). *J Cardiovasc Magn Reson*. 2020. <https://doi.org/10.1186/s12968-020-00631-1>.
38. Truong VT, Palmer C, Young M, Wolking S, Ngo TNM, Sheets B, et al. Right atrial deformation using cardiovascular magnetic resonance myocardial feature tracking compared with two-dimensional speckle tracking echocardiography in healthy volunteers. *Sci Rep*. 2020. <https://doi.org/10.1038/s41598-020-62105-9>.
39. Qu Y-Y, Buckert D, Ma G-S, Rasche V. Quantitative assessment of left and right atrial strains using cardiovascular magnetic resonance based tissue tracking. *Front Cardiovasc Med*. 2021. <https://doi.org/10.3389/fcvm.2021.690240>.

Publisher's Note

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

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

