



OPEN

# Assessment of deep learning segmentation for real-time free-breathing cardiac magnetic resonance imaging at rest and under exercise stress

Martin Schilling<sup>1,5</sup>, Christina Unterberg-Buchwald<sup>1,2,3,5</sup>, Joachim Lotz<sup>1</sup> & Martin Uecker<sup>1,2,4</sup>✉

In recent years, a variety of deep learning networks for cardiac MRI (CMR) segmentation have been developed and analyzed. However, nearly all of them are focused on cine CMR under breathhold. In this work, accuracy of deep learning methods is assessed for volumetric analysis (via segmentation) of the left ventricle in real-time free-breathing CMR at rest and under exercise stress. Data from healthy volunteers ( $n = 15$ ) for cine and real-time free-breathing CMR at rest and under exercise stress were analyzed retrospectively. Exercise stress was performed using an ergometer in the supine position. Segmentations of two deep learning methods, a commercially available technique (comDL) and an openly available network (nnU-Net), were compared to a reference model created via the manual correction of segmentations obtained with comDL. Segmentations of left ventricular endocardium (LV), left ventricular myocardium (MYO), and right ventricle (RV) are compared for both end-systolic and end-diastolic phases and analyzed with Dice's coefficient. The volumetric analysis includes the cardiac function parameters LV end-diastolic volume (EDV), LV end-systolic volume (ESV), and LV ejection fraction (EF), evaluated with respect to both absolute and relative differences. For cine CMR, nnU-Net and comDL achieve Dice's coefficients above 0.95 for LV and 0.9 for MYO, and RV. For real-time CMR, the accuracy of nnU-Net exceeds that of comDL overall. For real-time CMR at rest, nnU-Net achieves Dice's coefficients of 0.94 for LV, 0.89 for MYO, and 0.90 for RV and the mean absolute differences between nnU-Net and the reference are 2.9 mL for EDV, 3.5 mL for ESV, and 2.6% for EF. For real-time CMR under exercise stress, nnU-Net achieves Dice's coefficients of 0.92 for LV, 0.85 for MYO, and 0.83 for RV and the mean absolute differences between nnU-Net and reference are 11.4 mL for EDV, 2.9 mL for ESV, and 3.6% for EF. Deep learning methods designed or trained for cine CMR segmentation can perform well on real-time CMR. For real-time free-breathing CMR at rest, the performance of deep learning methods is comparable to inter-observer variability in cine CMR and is usable for fully automatic segmentation. For real-time CMR under exercise stress, the performance of nnU-Net could promise a higher degree of automation in the future.

The fast and reliable evaluation of cardiac function is an essential part of cardiac MRI (CMR), significant for patient diagnostics, disease analysis, therapy evaluation, follow-up, and risk estimation<sup>1,2</sup>. The main quantitative parameters of cardiac function are the left ventricular blood volume (LV), the volume of left ventricular myocardium (MYO), and the right ventricular blood volume (RV). These parameters are usually calculated by acquiring and segmenting a stack of cross-sectional images in short-axis view.

Advances in image reconstruction have enabled the continuous acquisition of high-quality images in real time, i.e. during free-breathing and independent of ECG-synchronization<sup>3-5</sup>. In CMR, real-time MRI has emerged as a

<sup>1</sup>Institute for Diagnostic and Interventional Radiology, Universitätsmedizin Göttingen, Göttingen, Germany. <sup>2</sup>German Centre for Cardiovascular Research (DZHK), Partner Site Göttingen, Göttingen, Germany. <sup>3</sup>Clinic of Cardiology and Pneumology, Universitätsmedizin Göttingen, Göttingen, Germany. <sup>4</sup>Institute of Biomedical Imaging, Graz University of Technology, Graz, Austria. <sup>5</sup>These authors contributed equally: Martin Schilling and Christina Unterberg-Buchwald. ✉email: uecker@tugraz.at

viable alternative for patients with arrhythmia<sup>6</sup>, for measurements using exercise stress<sup>7–9</sup>, as well as for real-time guidance in cardiac catheter interventions<sup>10–13</sup>.

The delineation of the LV boundary is an important step for the determination of end-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF). However, manual segmentation of images is tedious and affected by inter- and intra-observer variability<sup>14–16</sup>. In recent years, deep learning methods have been introduced into clinical practice for the generation of base contours, which are manually corrected as required. In research, a variety of deep learning methods have been developed for segmentation in CMR<sup>17</sup>. However, nearly all of these methods are focused on conventional breathhold cine imaging<sup>18</sup>. In real-time MRI, automatic segmentation becomes more important, because a series of heart beats is acquired instead of a single cine loop. Real-time exercise stress studies pose an additional challenge for automatic segmentation due to the tendency for inferior image quality<sup>8,19</sup> depending on an increased heart rate and breathing motion. Recent works covering real-time MRI<sup>20,21</sup> have used custom neural networks trained specifically for the application on real-time CMR, though have faced the problem of limited training data availability.

This study aims to investigate the feasibility of using deep learning methods trained on cine CMR data for the automatic segmentation of real-time free-breathing CMR images. It will evaluate the performance of the methods under the conditions of both rest and during exercise stress.

## Methods

### Overview

We analyzed cine and real-time measurements of healthy volunteers ( $n = 15$ ) acquired at rest and under exercise stress using a highly undersampled radial bSSFP sequence with NLINV reconstruction<sup>22,23</sup> with a temporal resolution of 33 ms.

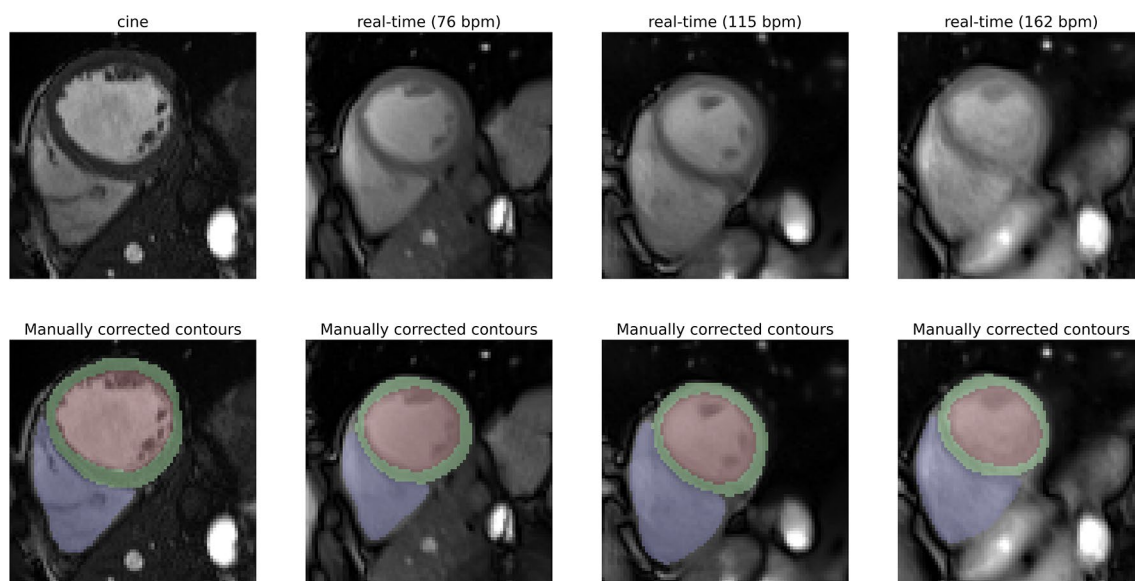
We evaluated segmentations obtained with two deep learning methods, (1) the automatic contour detection designed for cine CMR (comDL) that is included in the commercially available software Medis (version 4.0.56.4, QMass\*8.1, Medical Imaging Systems, Leiden, Netherlands) and (2) nnU-Net<sup>24</sup>, which was pre-trained on the cine CMR dataset of the Automated Cardiac Diagnosis Challenge (ACDC). Segmentations were compared to a reference model and standard clinical routine procedure of manually correcting the results obtained with comDL.

The accuracy of segmentation was evaluated for images in end-diastolic and end-systolic phases. The cardiac function parameters EDV, ESV, and EF were derived from neural network segmentation and compared to reference values derived from the manually corrected contours.

A comparison of cine and real-time CMR measurements and representative manually corrected contours are presented in Fig. 1.

### Dataset

The dataset consists of cine and real-time images from 15 healthy volunteers (7 male; 8 female; aged  $55 \pm 8$  (s.d.)) from a consecutive series of exams performed at the Institute for Diagnostic and Interventional Radiology of the University Medical Center Göttingen. It is part of a larger dataset acquired in a previous research study on



**Figure 1.** Comparison of different measurement types. Mid-ventricular slices in ED phase of the same volunteer for cine and real-time free-breathing at different heart rates (top) and corresponding manually corrected segmentation (bottom) in a short axis view are shown. Image quality decreases and reconstruction artifacts increase with an increasing heart rate. The left ventricular endocard (red), the left ventricular myocardium (green), and the right ventricle (blue) are segmented.

real-time MRI with the approval of the local ethics committee. Consent for publication was obtained from all participants in the study.

All volunteers were measured with the same protocol. The healthy volunteers underwent CMR in supine position using a 32-channel cardiac surface receiver coil at 3 T (Skyra, Siemens Healthineers, Germany). Conventional imaging at rest included a balanced steady-state free precession (bSSFP) ECG-gated cine sequence to create a short-axis stack covering the entire heart, including both ventricles and atria. Real-time CMR data acquisition was performed during free-breathing and without ECG-synchronization at rest and under two different levels of exercise stress. A detailed summary of acquisition parameters of cine and real-time CMR can be found in Supplementary Table S1 online.

Real-time acquisition was based on a bSSFP sequence using a highly undersampled radial encoding scheme and iterative image reconstruction<sup>3,23</sup>. Exercise stress CMR was performed using a CMR-compatible ergometer in the supine position (Lode, Leiden, Netherlands), as previously described in<sup>7</sup>.

After performing real-time free-breathing measurements at rest (RT), exercise stress was increased until a target heart rate of 110 beats per minute (bpm) was reached. Measurements at this target heart rate are referred to as real-time stress (RT stress). Finally, exercise stress was increased to the subjective, maximal exercise stress of each person and a measurement was performed (RT max stress). All persons gave written informed consent before each CMR examination.

## Segmentation

### Reference

To create a reference segmentation, manually corrected segmentation was performed using Medis (version 4.0.56.4, QMass<sup>®</sup>8.1, Medical Imaging Systems, Leiden, Netherlands). Contours for the left ventricular endocardium, the left ventricular epicardium, and the right ventricle were created on short axis slices in end-diastole and end-systole. The segmentation of the left ventricular myocardium is formed by the area between the left ventricular epicardium and endocardium. Contours automatically created by the software were then manually corrected to create the reference in accordance with the standard procedure used in clinical routine. Contour creation followed the contour protocol used for the Automated Cardiac Diagnosis Challenge (ACDC)<sup>14</sup>. Papillary muscles are included in the cavity and LV endocardial contours follow the limit defined by the aortic valve at the base of the LV.

For real-time measurements at rest and under physical stress, images in ED and ES phase in the whole time series (120–150 images) were segmented. At rest, the time series spans 3–4 heartbeats. Under physical stress, the time series spans 6–9 heartbeats. For real-time measurements at maximal physical stress, only images in ED and ES phase among the first 50 images of the time series (2–4 heartbeats) were segmented. For the measurements of three volunteers at maximal physical stress, image quality was too poor to create reasonable reference contours.

### Neural networks for automatic segmentation

This work evaluates two different neural networks for automatic segmentation. The first is the automatic contour creation from the commercial software Medis DL ACD (Medis deep learning automatic contour detection) in QMass 8.1. We refer to this method as comDL.

It should be stressed that comDL is intended to be used for cine CMR as a basis for the contours, to be manually corrected and checked as described above. Since manual correction is generally not feasible for the high number of images acquired using real-time MRI, the scope of this work is to evaluate the performance of deep-learning methods for automatic segmentation.

The second neural network evaluated in this work is nnU-Net. It is freely available and offers trained weights for various image segmentation challenges in the medical field. The model has already shown great versatility and was successfully used for a variety of medical segmentation tasks, e.g. achieving first place for all segmentation classes in the cardiac segmentation challenge “Automated Cardiac Segmentation Challenge” (ACDC)<sup>24</sup>.

Preliminary analysis (results not shown) showed that nnU-Net performs better if applied on individual 2D images rather than on a stack of 2D images, e.g. a time series of real-time MRI in a single slice or a stack of cine images within the same cardiac phase. The benefit of independent normalization of each image increased for more challenging segmentation tasks (see Supplementary Table S2 online). The application of 2D nnU-Net was identified as the best configuration for this task when compared with 3D nnU-Net and the ensemble of both models (see Supplementary Table S3 online).

Consequently, the 2D version of nnU-Net with weights pre-trained on the ACDC dataset was applied on single images of the dataset for all cine and real-time measurements.

In March 2023, a new version of nnU-Net was published (nnU-Net V2). As of 2024-01-19, no pre-trained weights have been published for the second version and segmentation performance reportedly remains the same<sup>25</sup>. As such, this work evaluates the first version of nnU-Net.

### Evaluation criterion

It evaluates the overlap of a predicted segmentation of a neural network  $P_k$  with a reference segmentation  $R_k$  for each individual segmentation class  $k$  and is defined as

$$DC = 2 \frac{P_k \cap R_k}{P_k + R_k}. \quad (1)$$

DC is a value between 0 and 1, with 0 denoting no overlap between prediction and reference and 1 denoting perfect agreement.

### Calculation of heart rates

We also categorized all real-time images based on heart rates, as these differed for RT max stress. For the calculation of heart rates, we used the three central slices of all slices between the base and apex. In the central slices, the left ventricle is present during the entire time series despite respiratory motion, which can cause a displacement of the heart in and out of the imaging plane in basal and apical slices. Heartbeats per minute (bpm) are calculated based on the duration between two end-diastolic phases.

### Cardiac function parameters

End-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF) were computed. To minimize the influence of respiratory motion on EDV and ESV, images in the ED and ES phase of the cardiac cycle during end-expiration were manually selected for each slice. For comDL and nnU-Net, cardiac function parameters were calculated based on segmentation of the same selected images. Ventricular volumes were calculated with Simpson's rule<sup>26,27</sup>.

For intra-observer variability, manually corrected contours for the derivation of the clinical measures of all volunteers were created three to six months after the initial segmentation. For inter-observer variability, manually corrected contours for the derivation of the clinical parameters were created for the first five volunteers by a second reader with experience in cardiac segmentation. Single images in the ED and ES phase during end-expiration were once again chosen from each slice and EDV, ESV, and EF were derived from newly created, manually corrected contours.

Previously reported<sup>15</sup> inter-observer variability was chosen as an additional reference for comparison because it represents the accuracy of the evaluation in the clinical workflow of cine CMR. Measurements are usually evaluated by a single person and the accuracy of the method is determined by the variance between different human observers.

### Statistics

We evaluated cardiac function parameters through Bland–Altman plots and paired two-sample t-tests. The cardiac function parameters EDV, ESV, and EF that were derived from comDL and nnU-Net segmentations were compared to the respective values derived from manually corrected contours (see Supplementary Figs. S1 and S2 online). The comparison includes cine CMR and real-time CMR measurements at rest and under stress of all volunteers. Additionally, cardiac function parameters of cine CMR and real-time CMR at rest were compared with each other (see Supplementary Fig. S3 online). T-tests were performed under the null hypothesis with a significance level of  $\alpha = 0.05$ .

## Results

### Segmentation accuracy

The data for all 15 volunteers could be successfully analyzed for cine CMR and real-time free-breathing CMR at rest and under exercise stress. For the measurements of three volunteers at maximal physical stress, image quality was too poor to create reasonable reference contours and thus only data of 12 volunteers were analyzed.

For cine CMR, the DC values for both neural network segmentations are comparable (Table 1) and show a high correlation with conventional segmentation. Figure 2 shows representative DC cases.

For cine CMR, the nnU-Net performance (Table 1) is comparable to the results achieved on the ACDC test dataset<sup>24</sup>. The Dice's coefficients reported here are slightly higher, as might be expected for the application on healthy subjects as compared to different pathology groups in the ACDC test dataset.

Real-time free-breathing measurements can be categorized based on the form of acquisition, e.g. at rest (RT), under a level of stress selected according to a targeted heart rate of 110 bpm (RT stress) and maximal exercise stress (RT max stress), for which the stress level and heart rate varies by volunteer.

Mean heart rate and standard deviation for all volunteers can be found as Supplementary Table S4 online. The different real-time free-breathing CMR measurements fall into different heart rate spans, RT—55–77 bpm, RT stress—107–120 bpm, RT max stress—121–164 bpm.

The accuracies of nnU-Net and comDL segmentation based on the heart rate categorization are presented in Table 1. The nnU-Net model shows good generalizability with high accuracies of segmentation for both RT (DC: LV 0.94, MYO 0.89, RV 0.90) and RT stress (DC: LV 0.92, MYO 0.85, RV 0.83). The accuracies for RT are in the order of previously reported<sup>15</sup> inter-observer variability for cine CMR. With the exception of the RV segmentation for RT, the accuracy of nnU-Net predictions exceeds the accuracy of comDL in real-time CMR.

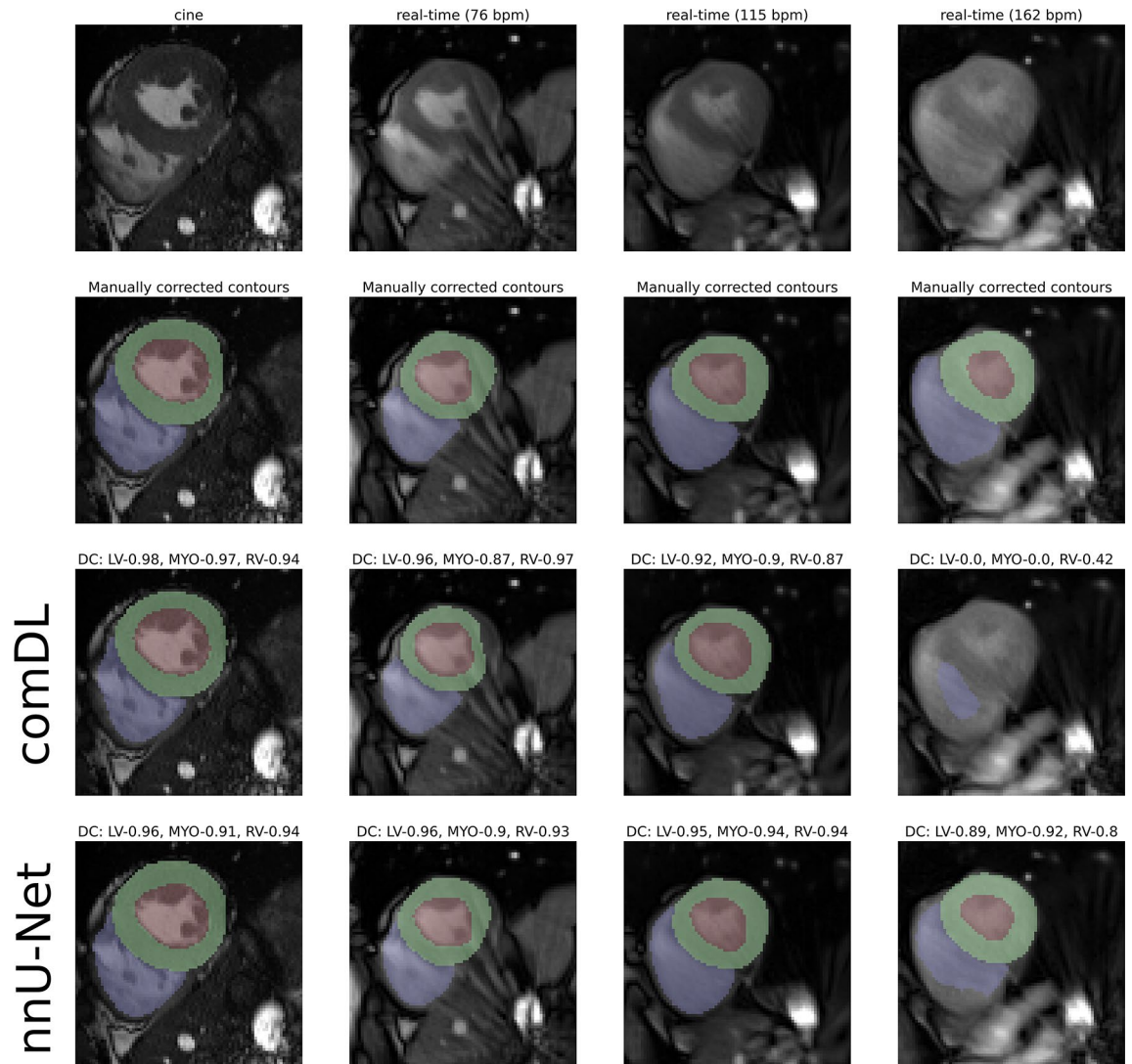
To better visualize the relationship between heart rate and the neural network segmentation accuracy, Fig. 3 shows the nnU-Net and comDL Dice's coefficients by calculated heart rate for real-time CMR. Two outliers are visible for the RV segmentation of the nnU-Net. Both data points correspond to the RT and RT stress measurements of the same volunteer. While the RV segmentation performs well for cine and RT it nearly completely fails for RT stress and RT maxstress, most likely because of smooth transitions between RV and neighboring tissue. A representative comparison can be found as Supplementary Fig. S4 online.

Example failure cases for the nnU-Net are shown in Fig. 4. Images in the basal and apical region of the heart were especially prone to segmentation failures.

### Cardiac function parameters

For cine and real-time CMR, the EDV, ESV, and EF were derived from the manually corrected contours for manually selected images in the ED and ES phases. The absolute values for each volunteer can be found in Supplementary Table S5 online. The comparison between these values and those derived from the nnU-Net and comDL segmentation is presented in Table 2 in form of absolute and relative differences.





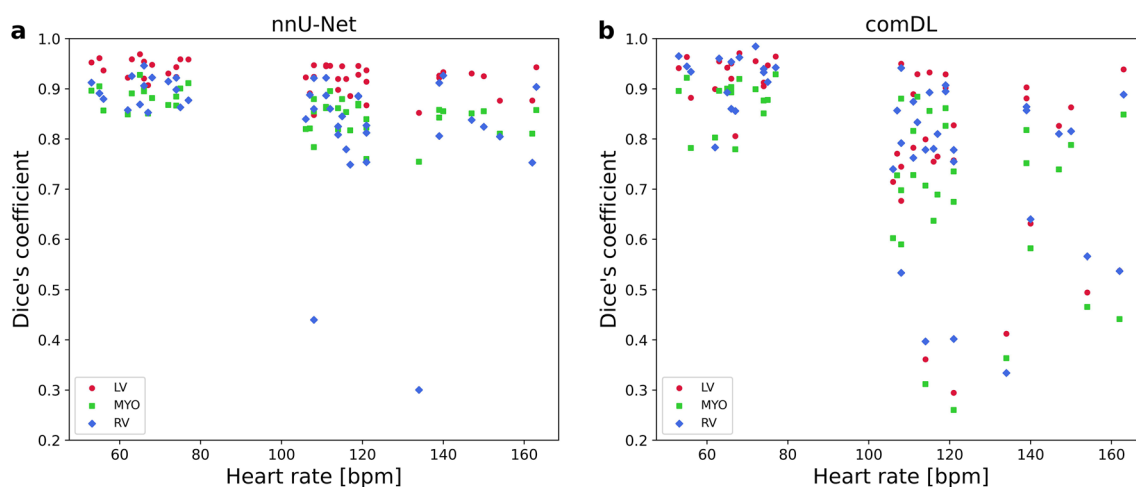
**Figure 2.** Representative segmentations of manually corrected contours and deep learning methods. Mid-ventricular slices in ES phase of a volunteer for cine and real-time free-breathing at different heart rates (first row) with corresponding manually corrected (second row), comDL (third row), and nnU-Net segmentation (fourth row). Accuracy of segmentation is measured with Dice's coefficient (DC). DC for left ventricular endocard (LV), left ventricular myocardium (MYO), and right ventricle (RV) are given for each segmentation.

For cine CMR, manual corrections of comDL led to only slight differences in EDV and ESV. For real-time CMR, the difference between the nnU-Net segmentation and manually corrected contours compares well to intra- and inter-observer variability for real-time CMR as well as to the inter-observer<sup>15</sup> variability previously reported for cine CMR (Table 2). For nnU-Net, mean relative differences of 2.4% for EDV, 6.6% for ESV, and 4.6% for EF were obtained for real-time CMR at rest. For nnU-Net, the absolute and relative differences of EDV, ESV, and EF are less than the inter-observer variability, although the direct comparability is limited because images were newly selected for intra- and inter-observer variability, while the nnU-Net segmentation has been compared to the manually corrected contours of the same images. The different selection process of the images may influence the variability a lot more than a different delineation of LV within the same image. Relative intra-observer variability increases from cine CMR to real-time CMR at rest, and again to real-time CMR at stress. Relative intra- and inter-observer variability is overall higher for ESV than for EDV, which is in agreement with previously reported inter-observer variability. Inter-observer variability is slightly higher for ESV in cine CMR than what was previously reported but only increases somewhat for RT and RT stress.

Clinical measures derived with nnU-Net and comDL are compared to references in form of Bland–Altman plots (Supplementary Figs. S1 and S2 online). According to paired two-sample t-tests, differences between nnU-Net and manually corrected contours are significant for ESV and EF for cine CMR ( $P < 0.001$ ) and not significant for EDV ( $P = 0.23$ ) for cine CMR and for all cardiac function parameters for RT and RT stress ( $P > 0.5$ ). Differences between comDL and manually corrected contours are not significant for EDV for cine ( $P = 0.16$ ) and real-time CMR ( $P = 0.61$ ) and significant for all other cardiac function parameters of cine, RT and RT stress ( $P < 0.05$ ). Individual  $P$  values are shown in the Bland–Altman plots.

n = 15	LV	MYO	RV
nnU-Net cine ED	0.98 (0.00)	0.91 (0.02)	0.92 (0.05)
nnU-Net cine ES	0.92 (0.04)	0.91 (0.03)	0.88 (0.06)
nnU-Net cine	0.95 (0.02)	0.91 (0.02)	0.90 (0.03)
nnU-Net RT	0.94 (0.02)	0.89 (0.02)	0.90 (0.03)
nnU-Net RT stress	0.92 (0.03)	0.85 (0.03)	0.83 (0.11)
nnU-Net RT max stress (n = 12)	0.91 (0.03)	0.83 (0.04)	0.79 (0.16)
comDL cine ED	0.98 (0.02)	0.97 (0.02)	0.92 (0.06)
comDL cine ES	0.95 (0.05)	0.95 (0.04)	0.88 (0.08)
comDL cine	0.97 (0.03)	0.96 (0.02)	0.90 (0.06)
comDL RT	0.93 (0.04)	0.88 (0.05)	0.92 (0.05)
comDL RT stress	0.79 (0.15)	0.72 (0.15)	0.79 (0.14)
comDL RT max stress (n = 12)	0.70 (0.21)	0.62 (0.19)	0.69 (0.18)
	0.94 (0.04)	0.88 (0.02)	0.87 (0.06)
Inter-observer cine (n = 50) <sup>15</sup>	0.92 (0.04)	0.87 (0.03)	0.88 (0.05)
	0.93 (0.04)	0.88 (0.02)	0.89 (0.05)

**Table 1.** Dice's coefficients of nnU-Net and comDL segmentation for cine and real-time CMR. The table features the mean and standard deviation (in parenthesis) of the Dice's coefficients for the left ventricular endocard (LV), the left ventricular myocardium (MYO), and the right ventricle (RV) for all volunteers. For cine CMR, images were separated into end-diastolic (ED) and end-systolic (ES) phase. For RT max stress, only data of 12 volunteers were analyzed, because in three cases the image quality was too poor to create reasonable reference contours. Mean and standard deviation (in parenthesis) of previously reported values for inter-observer variability from three different human observers for cine CMR are shown for comparison.

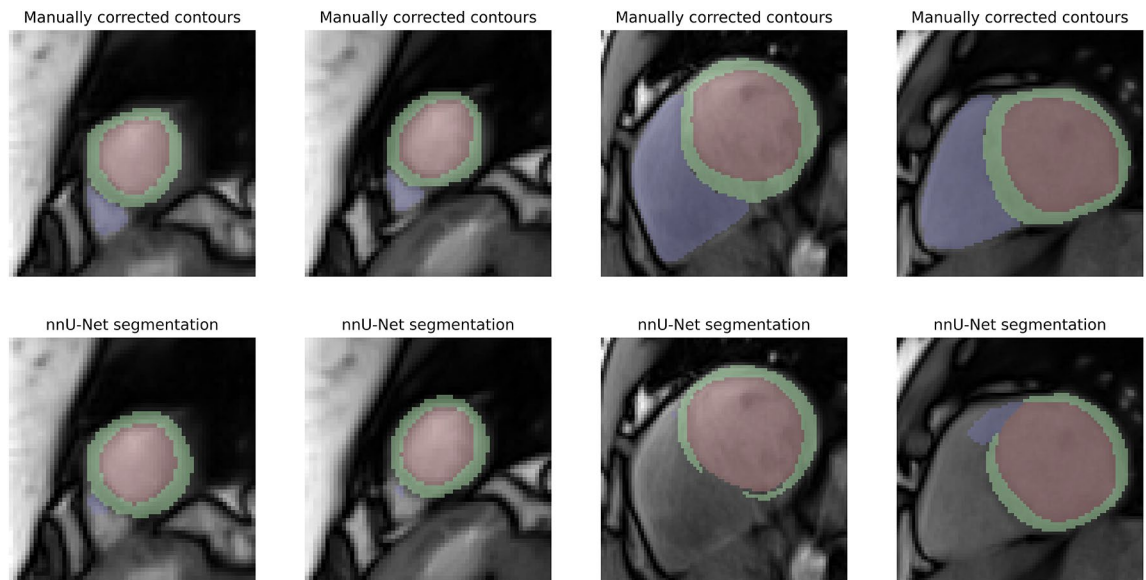


**Figure 3.** Dice's coefficient of real-time CMR measurements plotted against heart rate. DC values of LV, MYO, and RV are calculated for (a) nnU-Net and (b) comDL segmentation in respect to manually corrected contours. Each data point presents the average DC of a segmentation class for a single real-time measurement of a volunteer. Real-time CMR at rest, under exercise stress and maximal exercise stress are presented by their average calculated heart rate.

Although a comparison of cardiac function parameters between cine and real-time CMR was not the focus of this work, we note that the data show a higher mean EDV (mean difference 8.9 mL) and a lower mean ESV (mean difference -4.1 mL) for cine CMR (see Supplementary Fig. S3 online). These deviations are in the order of previously reported<sup>4</sup> differences between cine and real-time CMR. According to paired two-sample t-tests, differences between cine and real-time CMR are significant ( $P < 0.05$ ) for all cardiac function parameters for manually corrected contours.

## Discussion

In this study of cine CMR and real-time free-breathing CMR at rest and under exercise stress of 15 healthy volunteers, we found that the accuracy of comDL and nnU-Net is in the order of inter-observer variability for cine CMR and real-time CMR at rest. Consequently, both methods are viable for a prospective, automated evaluation



**Figure 4.** Example segmentation failures of nnU-Net for real-time CMR under exercise stress. Incomplete segmentation of the right ventricle in the apical region (first and second column). Anatomically incoherent segmentation of the myocardium and right ventricle in the basal region (third and fourth column).

with little or no manual correction. For real-time CMR under exercise stress, nnU-Net is usable as a basis for manually corrected contours.

In our study, comDL contours are the basis for manually corrected contours, which act as reference. These automatically created contours require little manual correction, confirming the good performance of comDL for cine CMR. However, our study showed comDL to be less accurate than nnU-Net for real-time CMR, especially for RT stress and RT max stress. These results might be expected from the perspective that comDL was designed and most likely optimized for cine CMR, for which it performed very well.

nnU-Net shows better generalizability for real-time CMR than comDL, having its segmentation accuracy decrease less between real-time and exercise stress measurements. For measurements under exercise stress, nnU-Net may have reached the limit of its applicability for fully automatic segmentation. Although its performance remains quite good, as demonstrated by high Dice's coefficients and a low mean absolute difference in ESV, the mean difference in EDV is significantly larger than intra-observer variability. While the accuracy of nnU-Net might not yet be sufficient for fully automatic segmentation, it shows promising results for an increased degree of automation in the future.

In this work, we observed the highest degree and frequency of deviations between reference and neural network segmentation in the basal and apical regions of the heart. These regions were also identified as the most problematic factor for neural network segmentation in the evaluation of ACDC<sup>14</sup>. Due to the ambiguity of slice positions, which can still include or exclude certain segmentation classes, comDL and nnU-Net showed some difficulty in correctly segmenting these areas. One approach to this issue is the usage of multiple neural networks individually trained on specific heart regions<sup>28</sup>. This however requires manually labeling input data prior to segmentation.

One previous study concerning the segmentation of real-time free-breathing CMR with deep learning neural networks is the work by Yang et al.<sup>20</sup>, which created a custom neural network based on the U-Net<sup>29</sup> architecture and also trained on the ACDC dataset. They evaluated their network on end-systolic and end-diastolic phases in the end-expiration state of healthy volunteers. Our results for real-time free-breathing CMR show a higher segmentation accuracy for comDL and nnU-Net compared to results of<sup>20</sup> (DC: LV 0.919, MYO 0.806, RV 0.818), showing the progress of segmentation networks for real-time CMR.

New methods have been developed for cardiac segmentation in recent years, e.g. the usage of transformers within neural networks for segmentation is a more recent idea than the use of convolutional neural networks and might be an essential element for future research<sup>30</sup>. A combination of nnU-Net with transformers<sup>31</sup> shows promising results, especially in an ensemble with the unmodified nnU-Net. However, the highest performance of deep learning neural networks for segmentation in the field of CMR is still achieved by convolutional neural networks, often based on the U-Net or the nnU-Net<sup>30</sup>. Therefore, it seems reasonable to have nnU-Net as the state-of-the-art method, in particular because of its accessibility through pre-trained weights.

The development and application of deep learning methods is an active research topic in radiology<sup>32–34</sup>. Standards for the reporting of artificial intelligence methods were established for the medical field as a whole in<sup>35</sup> and the topic was specifically discussed for CMR in<sup>36,37</sup>.

For cardiac segmentation, deep learning methods have already been successfully integrated into clinical routines and have substantially reduced the number of manual corrections needed. With recent progress, a fully automatic workflow seems feasible even for real-time CMR, but essential steps are still missing. If the evaluation of real-time free-breathing measurements should work analogously to cine CMR, images within the same

<b>(a) Absolute difference</b>			
<b>n = 15</b>	<b>EDV [mL]</b>	<b>ESV [mL]</b>	<b>EF [%]</b>
nnU-Net cine	1.8 (1.5)	3.7 (2.9)	2.5 (2.2)
nnU-Net RT	2.9 (2.1)	3.5 (3.6)	2.6 (2.6)
nnU-Net RT stress	11.4 (16.3)	2.9 (1.8)	3.6 (4.5)
comDL cine	1.0 (0.7)	3.0 (2.7)	1.8 (1.4)
comDL RT	6.0 (5.4)	5.4 (4.9)	4.0 (5.4)
comDL RT stress	32.7 (28.5)	7.3 (5.0)	26.3 (47.0)
Intra-observer cine	2.8 (3.9)	1.4 (1.9)	1.5 (1.6)
Intra-observer RT	3.5 (4.2)	4.5 (4.5)	2.9 (2.3)
Intra-observer RT stress	4.1 (3.3)	4.3 (3.9)	3.5 (2.2)
Inter-observer cine (n = 5)	4.4 (4.9)	8.5 (7.9)	4.5 (3.4)
Inter-observer RT (n = 5)	4.8 (3.1)	11.1 (7.6)	7.1 (4.8)
Inter-observer RT stress (n = 5)	11.5 (6.6)	9.4 (8.9)	5.1 (5.1)
	6.1 (4.4)	4.1 (4.2)	3.1 (2.1)
Inter-observer cine (n = 50) <sup>15</sup>	8.8 (4.8)	6.7 (4.2)	3.0 (2.4)
	4.8 (3.1)	7.1 (3.8)	3.8 (1.8)
<b>(b) Relative difference</b>			
<b>n = 15</b>	<b>EDV [%]</b>	<b>ESV [%]</b>	<b>EF [%]</b>
nnU-Net cine	1.2 (1.0)	6.7 (5.2)	4.2 (3.9)
nnU-Net RT	2.4 (2.4)	6.6 (10.1)	4.6 (4.2)
nnU-Net RT stress	7.2 (9.0)	5.5 (3.6)	5.9 (6.8)
comDL cine	0.7 (0.5)	5.2 (3.9)	2.9 (2.4)
comDL RT	4.7 (6.1)	8.1 (5.8)	7.3 (10.2)
comDL RT stress	22.8 (22.9)	14.2 (11.6)	42.1 (75.9)
Intra-observer cine	1.7 (2.3)	2.7 (3.8)	2.5 (2.7)
Intra-observer RT	2.4 (2.6)	6.3 (5.2)	5.2 (4.2)
Intra-observer RT stress	2.9 (2.5)	7.6 (5.4)	5.9 (4.1)
Inter-observer cine (n = 5)	3.2 (3.6)	15.9 (12.4)	7.1 (5.4)
Inter-observer RT (n = 5)	3.1 (1.8)	17.5 (10.5)	11.9 (8.5)
Inter-observer RT stress (n = 5)	8.3 (4.5)	16.9 (14.5)	9.2 (10.0)
	4.2 (3.1)	6.8 (7.5)	5.1 (3.7)
Inter-observer cine (n = 50) <sup>15</sup>	6.3 (3.3)	12.5 (8.5)	4.9 (3.8)
	3.4 (2.2)	11.7 (5.1)	6.6 (3.2)

**Table 2.** Cardiac function parameters from deep learning methods compared to references.. Cardiac function parameters derived from comDL and nnU-Net segmentation are compared to reference values obtained with manually corrected segmentation. The table features the differences of the cardiac function parameters of the left ventricular end-diastolic volume (EDV), the left ventricular end-systolic volume (ESV), and the left ventricular ejection fraction (EF) of cine and real-time CMR of all volunteers. The mean and standard deviation (in parenthesis) of the difference are reported for (a) the absolute and (b) the relative difference. Intra- and inter-observer variability are given for values which were derived from manually corrected contours of newly selected images. Previously reported values for inter-observer variability from three different human observers for cine CMR are presented for comparison.

respiratory motion state must be evaluated across slices. The identification of the respiratory state would need to be automated, e.g. with the help of an external device like a respiratory belt, or by automatic analysis of the images. Based on the respiratory motion state, images in the correct cardiac phase can then be automatically selected based on LV area. To keep manual corrections to a minimum, additional priors in the form of confidence maps, such as uncertainty maps<sup>38</sup>, could be used to quantify the need of manual correction. For patients with arrhythmia, arrhythmic heartbeats would need to be distinguished automatically from regular (sinus) rhythm.

Some limitations of our study must be noted. Firstly, extrasystolic heartbeats within the selected images were not excluded, as the limited duration of the time series did not always allow the monitoring of prior and subsequent heartbeats to fully exclude an extrasystole. However, this only affects the validity of the resulting cardiac function parameters, not the comparison between the manually corrected contours and the deep learning methods, as the same images have been evaluated for all methods. Secondly, only a relatively small number of healthy volunteers were included in this study and the validity for clinical application on patients still needs to be demonstrated. Thirdly, no detailed methodological information can be provided for comDL, as it is part of a commercial software. It still serves as a reference to the clinical standard. Finally, only two deep-learning methods have been compared, which limits the generalizability of the results to machine learning methods in general. We



hope to address this with the publication of image data and code to enable the testing and comparison of other deep-learning methods on the same data.

## Conclusion

In this study, we assessed the feasibility of automatic cardiac segmentation on real-time CMR using deep learning methods. Two deep learning methods originally designed or trained for segmentation of cine CMR were evaluated for cine and real-time MRI in comparison to a manually corrected reference segmentation. The segmentation accuracy is superior in cine CMR compared to real-time CMR at rest and diminishes further for real-time CMR under exercise stress. The accuracy for real-time CMR at rest is in the range of reported inter-observer variability of cine CMR for both networks. In this work, comDL shows very good performance for segmentation on cine CMR but less applicability for real-time CMR compared to nnU-Net. For real-time CMR at rest, cardiac function parameters obtained with nnU-Net segmentation are in the range of intra-observer variability. For real-time CMR under exercise stress, the performance of the deep learning methods - while still not sufficient for a fully automatic segmentation - is promising.

## Data availability

Cine and real-time CMR images, comDL and manually corrected contours, nnU-Net segmentation, and manually selected end-expiration indexes are available at Zenodo under <https://zenodo.org/records/10117943> (DOI:10.5281/zenodo.10117943).

## Code availability

Code for the reproduction of figures and results is available under <https://github.com/mrirecon/dl-segmentation-on-realtime-cmr>.

Received: 29 November 2023; Accepted: 9 February 2024

Published online: 14 February 2024

## References

- White, H. D. *et al.* Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction. *Circ. J.* **76**, 44–51. <https://doi.org/10.1155/2019/5636423> (1987).
- Norris, R. M., White, H. D., Cross, D. B., Wild, C. J. & Whitlock, R. M. L. Prognosis after recovery from myocardial infarction: The relative importance of cardiac dilatation and coronary stenoses. *Eur. Heart J.* **13**, 1611–1618. <https://doi.org/10.1155/2019/5636423> (1992).
- Uecker, M. *et al.* Real-time MRI at a resolution of 20 ms. *NMR Biomed.* **23**, 986–994. <https://doi.org/10.1002/nbm.1585> (2010).
- Feng, L. *et al.* Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE. *Magn. Reson. Med.* **70**, 64–74. <https://doi.org/10.1002/mrm.24440> (2013).
- Saybasili, H., Herzka, D. A., Seiberlich, N. & Griswold, M. A. Real-time imaging with radial grappa: Implementation on a heterogeneous architecture for low-latency reconstructions. *Magn. Reson. Imaging* **32**, 747–758. <https://doi.org/10.1016/j.mri.2014.02.022> (2014).
- Laubrock, K. *et al.* Imaging of arrhythmia: Real-time cardiac magnetic resonance imaging in atrial fibrillation. *Eur. J. Radiol. Open*. <https://doi.org/10.1016/j.ejro.2022.100404> (2022).
- Steinmetz, M. *et al.* Impaired exercise tolerance in repaired tetralogy of Fallot is associated with impaired biventricular contractile reserve: An exercise-stress real-time cardiovascular magnetic resonance study. *Circ. Cardiovasc. Imaging* <https://doi.org/10.1161/CIRCIMAGING.120.011823> (2021).
- Li, Y. Y. *et al.* Real-time exercise stress cardiac MRI with Fourier-series reconstruction from golden-angle radial data. *Magn. Reson. Imaging* **75**, 89–99. <https://doi.org/10.1016/j.mri.2020.10.010> (2021).
- Backhaus, S. J. *et al.* Exercise stress real-time cardiac magnetic resonance imaging for noninvasive characterization of heart failure with preserved ejection fraction. *Circ. J.* **143**, 1484–1498. <https://doi.org/10.1161/CIRCULATIONAHA.120.051542> (2021).
- Eitel, C., Hindricks, G., Grothoff, M., Gutberlet, M. & Sommer, P. Catheter ablation guided by real-time MRI. *Curr. Cardiol. Rep.* **16**, 1–7. <https://doi.org/10.1007/s11886-014-0511-6> (2014).
- Unterberg-Buchwald, C. *et al.* Targeted endomyocardial biopsy guided by real-time cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **19**, 45. <https://doi.org/10.1186/s12968-017-0357-3> (2017).
- Campbell-Washburn, A. *et al.* Real-time MRI guidance of cardiac interventions. *J. Magn. Reson. Imaging* **46**, 935–950. <https://doi.org/10.1002/jmri.25749> (2017).
- Franson, D., Dupuis, A., Gulani, V., Griswold, M. & Seiberlich, N. A system for real-time, online mixed-reality visualization of cardiac magnetic resonance images. *J. Imaging* <https://doi.org/10.3390/jimaging7120274> (2021).
- Bernard, O. *et al.* Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?. *IEEE Trans. Med. Imaging* **37**, 2514–2525. <https://doi.org/10.1109/TMI.2018.2837502> (2018).
- Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **20**, 65. <https://doi.org/10.1186/s12968-018-0471-x> (2018).
- Bhuva, A. N. *et al.* A multicenter, scan-rescan, human and machine learning CMR study to test generalizability and precision in imaging biomarker analysis. *Circul.: Cardiovasc. Imaging* **13**, 684–695. <https://doi.org/10.1016/j.jcmg.2019.05.030> (2019).
- Chen, C. *et al.* Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* <https://doi.org/10.3389/fcvm.2020.00025> (2020).
- Shoib, M. A. *et al.* An overview of deep learning methods for left ventricle segmentation. *Comput. Intell. Neurosci.* <https://doi.org/10.1155/2023/4208231> (2023).
- Morales, M. A. *et al.* Highly accelerated free-breathing real-time myocardial tagging for exercise cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **25**, 56. <https://doi.org/10.1186/s12968-023-00961-w> (2023).
- Yang, F. *et al.* A deep learning segmentation approach in free-breathing real-time cardiac magnetic resonance imaging. *Biomed. Res. Int.* <https://doi.org/10.1155/2019/5636423> (2019).
- Qi, Y., Wang, F., Cao, J. J. & Li, Y. Y. A deep learning approach to real-time volumetric measurements without image reconstruction for cardiovascular magnetic resonance. *Physiol. Meas.* <https://doi.org/10.1088/1361-6579/ac92bc> (2022).
- Uecker, M., Zhang, S. & Frahm, J. Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magn. Reson. Med.* **63**, 1456–1462. <https://doi.org/10.1002/mrm.22453> (2010).

23. Zhang, S., Uecker, M., Voit, D., Merboldt, K. & Frahm, J. Real-time cardiovascular magnetic resonance at high temporal resolution: Radial FLASH with nonlinear inverse reconstruction. *J. Cardiovasc. Magn. Reson.* **12**, 39 (2010).
24. Isensee, F., Jäger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211. <https://doi.org/10.1038/s41592-020-01008-z> (2021).
25. Isensee, F. Welcome to the new nnu-net! <https://github.com/MIC-DKFZ/nnUNet/tree/master> (2023). Accessed 26 Oct 2023.
26. Greenberg, S. Assessment of cardiac function: Magnetic resonance and computed tomography. *J. Thorac. Imaging* **15**, 243–251. <https://doi.org/10.1097/00005382-200010000-00004> (2000).
27. Mahnken, A., Günther, R. & Krombach, G. The basics of left ventricular functional analysis with MRI and MSCT. *Rofo-Fortschr. Röntg.* **176**, 1365–1379. <https://doi.org/10.1055/s-2004-813562> (2004).
28. Amirrajab, S., Al Khalil, Y., Pluim, J., Breeuwer, M. & Scannell, C. M. Cardiac MR image segmentation and quality control in the presence of respiratory motion artifacts using simulated data. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers* (eds Camara, O. et al.) 466–475 (Springer, 2022). [https://doi.org/10.1007/978-3-031-23443-9\\_44](https://doi.org/10.1007/978-3-031-23443-9_44).
29. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (eds Navab, N. et al.) 234–241 (Springer, 2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
30. El-Taraboulsi, J., Cabrera, C. P., Roney, C. & Aung, N. Deep neural network architectures for cardiac image segmentation. *Artif. Intell. Life Sci.* **4**, 100083. <https://doi.org/10.1016/j.aills.2023.100083> (2023).
31. Zhou, H.-Y. et al. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Trans. Image Processing* **32**, 4036–4045. <https://doi.org/10.1109/TIP.2023.3293771> (2023).
32. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. W. L. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510. <https://doi.org/10.1038/s41568-018-0016-5> (2018).
33. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence—An ESR white paper. *Insights Imaging* **10**, 500–510. <https://doi.org/10.1186/s13244-019-0738-2> (2019).
34. Reardon, S. Rise of robot radiologists. *Nature* **576**, 54–58. <https://doi.org/10.1038/d41586-019-03847-z> (2019).
35. Mongan, J., Moy, L. & Kahn, C. E. Checklist for artificial intelligence in medical imaging (claim): A guide for authors and reviewers. *Radiol. Artif. Intell.* **2**, e200029. <https://doi.org/10.1148/ryai.2020200029> (2020).
36. Alabed, S. et al. Quality of reporting in AI cardiac MRI segmentation studies—A systematic review and recommendations for future studies. *Front. Cardiovasc. Med.* <https://doi.org/10.3389/fcvm.2022.956811> (2022).
37. Maiter, A., Salehi, M., Swift, A. J. & Alabed, S. How should studies using AI be reported? Lessons from a systematic review in cardiac MRI. *Front. Radiol.* <https://doi.org/10.3389/fradi.2023.1112841> (2023).
38. Sander, J., de Vos, B. D., Wolterink, J. M. & Išgum, I. Towards increased trustworthiness of deep learning segmentation methods on cardiac MRI. In *Medical Imaging 2019: Image Processing* Vol. 10949 (eds Angelini, E. D. & Landman, B. A.) 1094919 (SPIE, 2019). <https://doi.org/10.1117/12.2511699>.

## Acknowledgements

We are particularly grateful for the assistance given by Florian Kohler who recruited volunteers and performed the scans together with Ulrike Köchermann and Tanja Otto.

## Author contributions

M.S. experimental analysis, interpretation of data, and writing of manuscript; C.U.B. design of the MRI study, experimental workup, volunteer recruitment, and revision of the manuscript; J.L. revision of the manuscript; M.U. development of real-time MRI methods, data analysis, and revision of the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

MU is a co-inventor of a patent covering the real-time MRI technique used in this study. All other authors hereby state that they have no financial or personal relationships with other people or organizations that could inappropriately influence or bias their work.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-54164-z>.

**Correspondence** and requests for materials should be addressed to M.U.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024