Open-Source Pre-Clinical Image Segmentation: Mouse cardiac MRI datasets with a deep learning segmentation framework

AUTHORS:

Wan Nur Hanisah [a,b], Daniel Stuckey [b], Tina Yao [a], Mark Wrobel [a], Ruaraidh Campbell [a], Vivek Muthurangu [a], Jennifer Steeden [a*]

a. UCL Centre for Translational Cardiovascular Imaging, Institute of Cardiovascular Science, University College London, London, UK
b. UCL Centre for Advanced Biomedical Imaging, University College London, London, UK

CORRESPONDING AUTHOR:

Jennifer Steeden

jennifer.steeden@ucl.ac.uk

KEY WORDS:

Pre-clinical, Segmentation, Deep Learning, Open-Source

ABSTRACT:

Longitudinal cardiac MRI in small animal models is essential for studying cardiovascular disease mechanisms and assessing new therapies. However, conventional manual segmentation of ventricular structures from cine short-axis images is slow, labour-intensive, and subject to substantial observer variability, limiting reproducibility in large-scale pre-clinical studies. While deep learning (DL) segmentation has become standard in human cardiac MRI, existing models do not generalize to pre-clinical datasets. Current mouse-specific approaches have been constrained by small private cohorts, where the DL segmentation models are not published.

Here, we present the first publicly-available pre-clinical cardiac MRI dataset, along with an open-source DL segmentation model and a web-based interface for easy deployment.

The dataset comprises of complete cine short-axis cardiac MRI images from 130 mice with diverse phenotypes. The dataset also contains expert manual segmentations of left ventricular (LV) blood pool and myocardium at end-diastole, end-systole, as well as additional timeframes with artefacts to improve robustness.

Using this resource, we developed an open-source DL segmentation model based on the UNet3+ architecture. The model was evaluated against an independent test data (n=25 mice) and compared to expert segmentations in terms of DICE and functional parameters. At inference the model took <20 ms per 2D image, enabling complete cine stack segmentation in ~4.6 s — over 6,000-fold speed-up over manual analysis. It achieved high segmentation accuracy in both the blood pool and myocardium (Dice coefficient mean ± standard deviation: 0.94 ± 0.02 and 0.91 ± 0.03, respectively). In addition, functional parameters calculated from manual ground-truth and DL segmentations demonstrated excellent agreement, with small, non-significant biases (end-diastolic volume r=0.94; end-systolic volume r=0.99; ejection fraction r=0.98; stroke volume r=0.92; myocardial mass r = 0.96).

In summary, we provide the first open-access mouse cardiac MRI dataset, an open-source DL segmentation model, and an accessible inference platform. These tools establish a benchmark for pre-clinical cardiac MRI, enabling reproducible, scalable, and community-driven development of DL methods to accelerate cardiovascular research.

INTRODUCTION

Longitudinal monitoring of cardiovascular function in pre-clinical small animal models is essential for understanding disease progression and evaluating new therapeutic strategies. Cardiac MRI has become the gold-standard method for assessing cardiac function, allowing non-invasive quantification of ventricular volumes, ejection fraction and myocardial mass. These metrics are conventionally derived from segmentation of 2D multi-slice short-axis (SAX) cine images, at end-diastole (ED) and end-systole (ES). However, current manual segmentation approaches are time-consuming and labour-intensive, which is particularly problematic when performing therapeutic trials with large numbers of animals. In addition, manual segmentations are prone to high levels of intra- and inter-observer variability [1-3], which limits reproducibility. Therefore, it is necessary to develop rapid and reproducible solutions for analysis of large, longitudinal pre-clinical studies.

In human studies, many tools have been developed to automate segmentation of cardiac MRI data and improve reproducibility, with current state-of-the-art methods relying on deep learning (DL) [4-6]. These methods have been shown to be rapid and accurate, as well as being robust to multi-site and multi-vendor data in large cohort studies [7, 8]. However, DL segmentation models trained using human cardiac MRI data perform poorly when applied to pre-clinical cardiac MRI images.

Pre-clinical cardiovascular research commonly uses rodent models, with recent papers showing DL models for segmentation of the ventricles from cardiac MRI [9-11]. In rats, these DL models have been trained with large amounts of data (from > 150 rats) and can achieve high segmentation accuracy, with low bias in clinical metrics and lower variability than seen between human observers. However, mice are the most widely used species for animal studies of cardiovascular disease [12], as they have a short gestation time, low maintenance/housing costs and can be genetically modified to enable the investigation into mechanisms underlying pathogenesis. Current DL models for segmentation of mouse cardiac MRI data have been trained with low numbers of datasets (from < 21 mice) [13-

15], limiting their accuracy and generalizability. This is problematic as mouse images have inherently lower signal-to-noise (SNR) and more artefacts, due to higher heart rates and smaller heart size of mice. In addition, all current pre-clinical DL segmentation methods rely exclusively on private datasets for training, making comparison of new methodologies difficult, and the current DL segmentation models are not published, meaning they cannot be used by the wider scientific community. To truly harness the full potential of DL segmentation models in mice, large publicly-available datasets, as well as open-source DL models are needed.

Therefore, this study aims to: (i) Create the first publicly-available database of pre-clinical mouse cardiac MR images, with expert contours; (ii) Develop an open-source DL segmentation model, optimized for left ventricular (LV) blood pool and myocardial segmentation in this pre-clinical mouse dataset; (iii) Validate the open-source DL model in terms of segmentation accuracy and conventional ventricular volumes in mice; (iv) Provide a web-based interface for easy deployment of this segmentation model.

METHODS

Publicly-Available Pre-clinical Cardiac MRI Dataset

The dataset includes 130 mice, imaged between June 2012 and December 2019. Each mouse consists of a complete short-axis cine stack, with ~9 slices (range 8–11) and ~22 cardiac timeframes (range 17–36), resulting in a total of 26,230 2D images. The dataset encompasses a broad range of myocardial phenotypes: (i) Controls (n = 39); (ii) Myocardial infarction with grafting of a biomaterial patch (n = 11); (iii) Transgenic model of dilated cardiomyopathy (n = 51); and (iv) Transverse aortic constriction model (n = 29).

All experiments were conducted in full compliance with the European Union Directive 2010/63/European Union and the UK Animals Scientific Procedure Act 1986. Experimental protocols were approved by the UK Home Office under Project Licenses

70_8709 and PP1692884, and by the Animal Welfare and Ethical Review Body at University College London (UCL).

Imaging Protocol

All imaging was performed on a 9.4 T small-animal MRI system (Agilent Technologies, Santa Clara, USA) equipped with 1000 mT/m gradient inserts and a 39 mm volume resonator RF coil (RAPID Biomedical, Rimpar, Germany). A small animal physiological monitoring system (SA Instruments, Stony Brook, NY) was used to maintain depth of anaesthesia and animal physiology. The electrocardiogram (ECG) trace was recorded using 3-lead subcutaneous electrodes, respiration rate was measured by a pressure sensitive balloon and internal temperature by rectal thermometer. Animals were anesthetised under a mixture of 1–2% isoflurane in oxygen.

For each mouse, cardiac-gated, spoiled gradient echo (GRE) cine MRI was acquired as described in [16]. A stack of 8-11 contiguous slices was used to ensure full coverage of the left ventricle. Prospective ECG triggering was used to synchronize image acquisition with the cardiac cycle, and respiratory gating was used to remove breathing motion. The imaging parameters were: slice thickness = 1.0 mm; TE/TR = 1.2/5.0 ms; flip angle = ~16°; acquisition matrix size = 128 × 128; interpolation factor = 2.0; interpolated pixel size = 0.1 × 0.1 mm. The temporal resolution was 5.0 ms, resulting in 17-36 timeframes per cardiac cycle (depending on heart rate).

Ground-truth Segmentation

Manual segmentation of the LV blood pool and myocardium was performed on all slices, at end-diastole and end-systole by one observer (WNH, 2 years' experience) using Horos version 3.3.6 (The Horos Project, https://horosproject.org). Trabeculae and papillary muscles were included in the myocardial mass. The contours were checked by two independent experts (DS and VM, 22 and 23 years' experience respectively) who had to reach consensus in case of discordance.

It was observed that the cine images often contain inflow-related artifacts in early-diastolic and mid-systolic timeframes. Therefore, to improve model robustness and enable use of the DL segmentation model throughout the cardiac cycle, additional timeframes containing inflow artefact were also segmented. Specifically, segmentations were performed in all slices for 44 additional cardiac timeframes (taken from four mice). This resulted in datasets from 304 cardiac timeframes (each containing all slices of the SAX), i.e. 2,883 2D images, with expert LV blood pool and myocardial segmentations.

Data Format and Availability

The dataset is shared in HDF5 format, with one .h5 file per mouse. Each HDF5 file contains the full cine SAX dataset and the expert segmentations for the LV blood pool and myocardium. For each mouse the imaging data (h5 dataset name *'Images'*) is of size: *matrix_size_x* × *matrix_size_y* × *number_of_slices* × *number_of_time_frames*, where the matrix size is 256 × 256, the number of slices is in the range 8-11 and the number of timeframes is in the range 17-36. The segmentation data (h5 dataset name *'Masks'*) is the same size as the image data, with values in the range 0-2, representing voxels belonging to the background (0), the LV myocardium (1) and LV blood pool (2). A third parameter is given in the HDF5 file which identifies the cardiac timeframes which contain expert segmentation (h5 dataset name *'Frames_with_ROI'*). Cardiac timeframes which do not contain expert segmentation consist of only zeros in the '*Masks'* data.

The dataset for all 130 mice, is available at: https://github.com/mrphys/Open-Source_Pre-Clinical_Segmentation.git

Open-Source Pre-clinical Segmentation Model

A number of network architectures have been demonstrated for medical image segmentation, including the UNet [17], nnUNet [18] and TransUNet [19]. In this study we developed an open-source DL segmentation model based on the UNet3+ architecture

[20]. This is a network architecture which we have previously shown to achieve very good segmentation quality in complex congenital heart disease in humans [21]. The UNet3+ model consists of an advanced encoder-decoder network with deep supervision, designed to enhance multi-scale feature integration, shown in Figure 1. The network features two layers per block, five scales, and batch normalization between layers. Final predicted labels are obtained by assigning each pixel to the class with the highest probability.

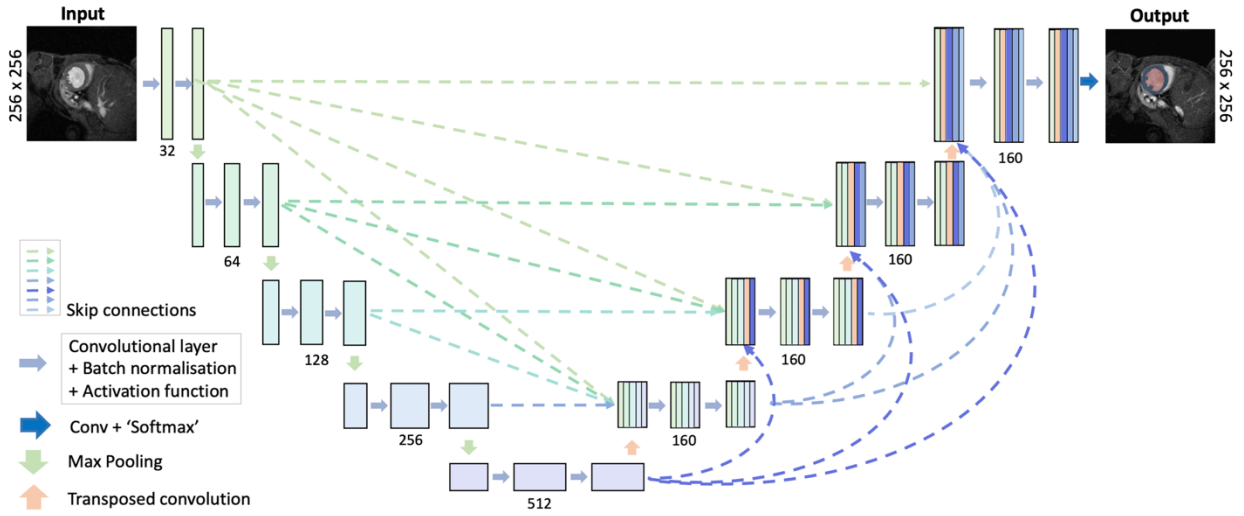

Figure 1: UNet3+ structure used in our open-source deep learning model. The DL model provides fully automated segmentation of LV blood pool and myocardial. The numbers at each level represent the number of channels.

To reduce the risk of overfitting and improve model generalizability, we applied on-the-fly data augmentation to each 2D image-mask pair during training, using the python Albumentations library [22]. Augmentation included: (i) Horizontal flips (p = 0.5); (ii) 90° rotations (p = 0.5); (iii) Affine transformations with controlled perturbations in scale (±10%), rotation (±10°), and shift (±6.25%); (iv) Brightness adjustments (limit = 0.17); (v) Gamma adjustments (range = 70 to 120); (vi) Contrast-limited adaptive histogram equalization (CLAHE, clip limit = 2.0); and (vii) Gaussian noise (variance range = 30 to 100).

We trained the model with an early stopping criterion (defined as no decrease in validation loss for 25 consecutive epochs), using the Adam optimizer with a batch size of 16 and a Dice loss. All training was performed using TensorFlow on a Linux Workstation (with NVIDIA GeForce RTX 3090).

Our publicly-available dataset was split into 90/15/25 mice for training/validation/testing, respectively. The timeframes which contained ground-truth segmentations were extracted, resulting in 224/30/50 SAX stack image-mask pairs, respectively. All timeframes which contained inflow-related artifacts were placed in the training dataset. All 2D slices from these SAX stacks were extracted as inputs to the model, resulting in 2087/286/460 2D image-mask pairs for training/validation/testing. Slices beyond the base/apex of the heart that did not contain any blood pool or myocardial segmentations were labelled as background. The images were normalized to standardize intensity values (within the 0 to 1 range) before inputting to the model, with one-hot encoding used to convert the masks to categorical data with three channels representing; background, LV myocardium and LV blood pool.

The model training code and final weights are available at: https://github.com/mrphys/Open-Source_Pre-Clinical_Segmentation.git

DL Segmentation Model Evaluation

The DL segmentation model was evaluated on the test dataset against the expert segmentations (460 2D images, taken from all slices of the SAX stack at ED and ES, of the 25 test mice). Prior to evaluation, only the largest connected component were kept, as identified using connected component labelling and centroid-based tracking, for both blood pool and myocardial classes. This post-processing step was used to remove misclassified ‘islands’ from the segmentations.

The accuracy of segmentation was quantified using Dice score. As well as overall Dice, we computed per-class Dice and ED and ES Dice, allowing assessment of interaction between Dice errors in cardiac structure and cardiac phase.

Additionally, DL-derived end-diastolic volume (EDV) and end-systolic volume (ESV) were calculated per mouse (n = 25) by summing the labelled blood pool voxels (scaled by voxel volume) across all slices from ED and ES timeframes, respectively. Ejection fraction (EF) was calculated as (EDV-ESV)/EDV, and stroke volume (SV) as EDV-ESV. Myocardial mass was calculated at ED and ES by calculating the sum of the labelled myocardial voxels (scaled by voxel volume and myocardial density – 1.05 g/ml). We also investigated the relationship between Dice score and errors in clinical metrics (EDV, ESV, EF, SV and myocardial mass).

Web-Based Inference

We have developed an open-source web-based application, to enable an easy-to-use interface to our DL segmentation model, without the need for local installation. Firstly, an interactive web pipeline was built using *Streamlit*, an open-source Python framework (https://github.com/streamlit/streamlit). The *Streamlit* application is hosted on *Hugging Face* – an open-source platform and ecosystem for building, sharing, and using machine learning models [23]. *Hugging Face* was chosen as their *Spaces* platform provides basic usage with no charge (on a vCPU with 16 GB RAM). The data is encrypted in transit using industry-standard encryption (via TLS/SSL). Uploaded data are processed in-memory (with temporary files deleted after use), and no input data or model outputs are stored.

The *Streamlit* application for our DL segmentation model requires the complete SAX cine image data to be uploaded in NIfTI format, as a zip file using a simple file browser. Pre-processing and inference are performed on all 2D images. The resulting blood-pool and

myocardial volumes are combined across all slices at each timeframe and output in a .csv file. The blood-pool volumes are used to easily identify ED and ES, and these volumes are displayed as a GIF, with the segmentations overlaid.

The web-based inference is available at: https://huggingface.co/spaces/mrphys/Pre-clinical_DL_segmentation

Statistics

Statistical analyses were performed using the SciPy library (version 1.9.1) in Python, with a *P-value* less than .05 considered statistically significant. All results are expressed as the mean ± standard deviation. Dice scores between ground-truth segmentations and DL segmentations were compared using a paired t-test. A repeated measures ANOVA was used to evaluate differences in Dice scores across cardiac timeframes (ED vs ES) and anatomical structures (myocardium vs blood pool). Ventricular volumes (EDV and ESV), function (EF and SV) and myocardial mass (at ED and ES) measured using the expert ground-truth segmentations and DL segmentations were compared using a paired t-test. For assessment of agreement of clinical metrics (ventricular volumes, function and mass) the ground-truth segmentations were used as the clinical standard for Bland-Altman analysis. Association of the differences between the ground-truth segmentations and DL segmentations for clinical metrics were tested using Spearman's rank correlation coefficient. The relationship between Dice score and errors in clinical metrics were tested using Spearman's rank correlation coefficient.

RESULTS

Training of our open-source DL segmentation model took ~4 hours (early stopping at 163 epochs). At local inference the model took 19.6 ± 1.9 ms per 2D image (range: 17.2 – 21.8 ms), resulting in 369.4 ± 16.6 ms for all SAX slices at ED and ES. Conventional manual segmentation for all SAX slices at ED and ES takes ~40 minutes, therefore the

DL segmentation represents ~x6,500 speed-up. Furthermore, our model can perform segmentation of the complete cine stack (all slices of all timeframes) in 4620 ± 676 ms, compared to approximately > 6 hours if done manually.

Segmentation Accuracy

Our open-source DL segmentation model showed strong agreement with ground-truth segmentations, with the best/median/worse Dice scores (for combined blood pool and myocardium, across all slices of one timeframe of 0.96 / 0.93 / 0.87, respectively, as shown in Figure 2 (corresponding movies in Supplementary Information 1-3).

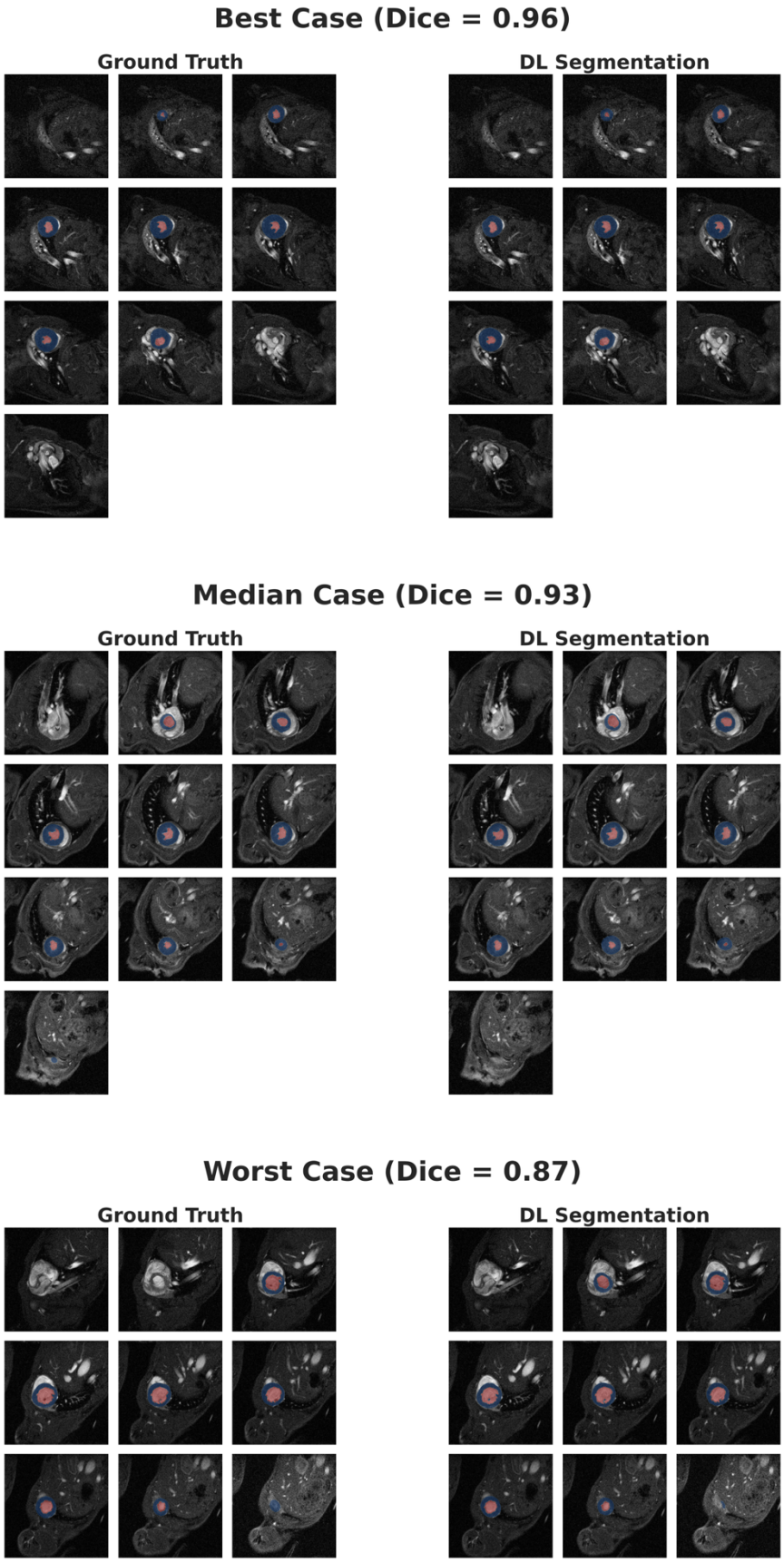

Figure 2: Ground-truth versus deep learning segmentations for the best, median, and worst overall Dice score in the test dataset. The images show all slices from base-to-apex, with the blood pool segmentation shown in red and the myocardial segmentation in blue. Corresponding movies can be found in Supplementary Information S1-S3.

It should be noted that there was a small but statistically significant difference between blood pool and myocardial accuracy ($p < 0.05$) and a significant interaction between cardiac phase and structure ($p < 0.05$). Specifically, the DL segmentation model was more accurate for the blood pool than the myocardium (Dice: 0.94 ± 0.02 vs 0.91 ± 0.03, respectively, $p < 0.001$). In addition, blood pool segmentations were more accurate in ED than ES (Dice: 0.95 ± 0.02 vs 0.93 ± 0.03, respectively, $p < 0.01$), while myocardial segmentations were better in ES than ED (Dice: 0.92 ± 0.02 vs 0.89 ± 0.03, respectively, $p < 0.002$).

Cardiac Function Accuracy

Clinical cardiac volume and function metrics measured from ground-truth segmentations and from the DL segmentations are shown in Table 1. Bland-Altman and correlation plots for are shown in Figure 3. There was good agreement in EDV ($r = 0.94$), with a small (+0.6 μL) overestimation from the DL segmentation, which did not reach statistical significance ($p = 0.36$). Additionally, there was good agreement in ESV ($r = 0.99$), with a small (+0.4 μL) overestimation from the DL segmentation, which also did not reach statistical significance ($p = 0.33$). SV demonstrated strong agreement ($r = 0.92$), with a small overestimation which did not reach significance (+0.3 μL, $p = 0.70$). A strong agreement ($r = 0.99$) was also observed in EF, with a small underestimation which did not reach significance (-0.2 %, $p = 0.70$). The model slightly overestimated myocardial mass at both ED and ES (2.3 mg and 1.6 ms, respectively), although neither difference was statistically significant ($p > 0.16$).

	Ground-truth segmentation	DL segmentation	Bias (Limits of Agreement)	p-value
EDV (μL)	65.6 ± 10.0	66.2 ± 10.1	0.6 (-6.0 to 7.1)	0.365
ESV (μL)	27.3 ± 10.1	27.7 ± 10.7	0.4 (-3.2 to 3.9)	0.330
SV (μL)	38.3 ± 6.3	38.5 ± 8.1	0.3 (-6.4 to 6.9)	0.696
EF (%)	59.3 ±10.7	59.1 ± 12.3	-0.2 (-5.7 to 5.3)	0.698
Myocardial Mass at ED (mg)	105.2 ± 26.8	107.5 ± 25.5	2.3 (-12.8 to 17.4)	0.165
Myocardial Mass at ES (mg)	111.2 ± 15.6	112.8 ± 26.1	1.6 (-13.0 to 16.3)	0.299

Table 1: Clinical metrics extracted from expert ground-truth segmentations and DL segmentations.

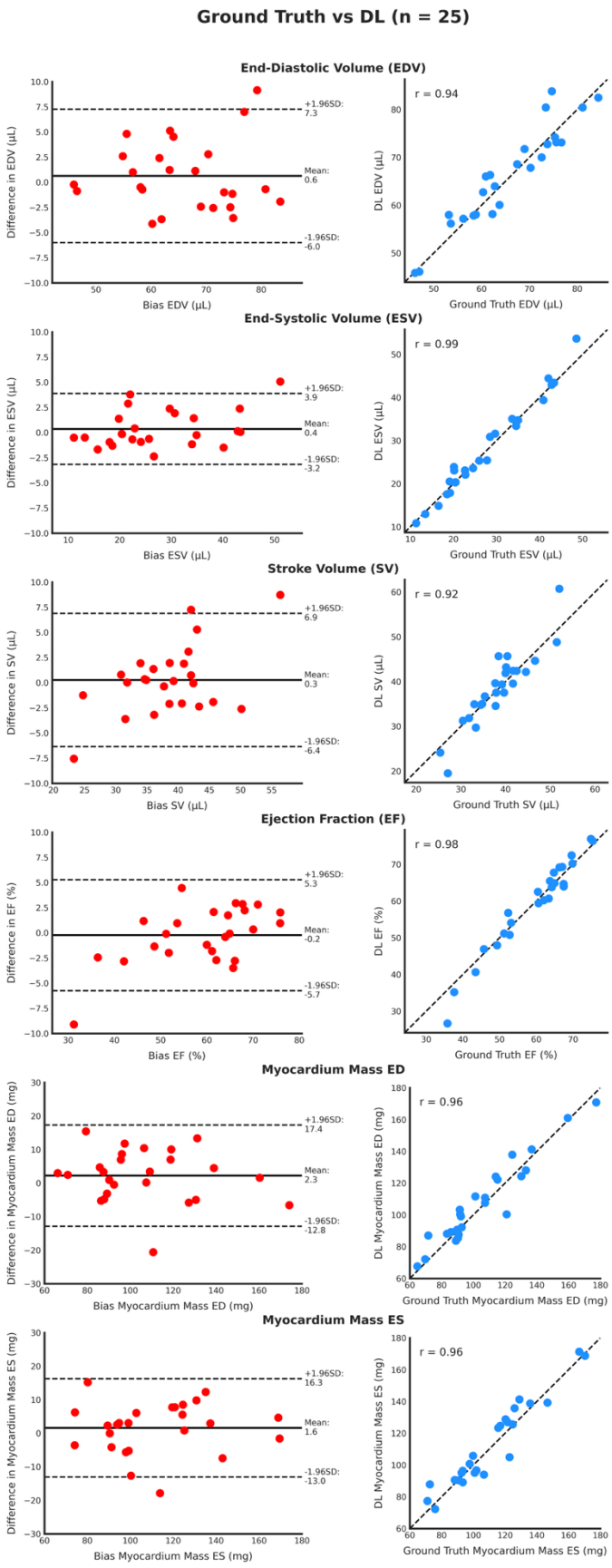

Figure 3: Bland-Altman plots (left) and line of equality plots (right) comparing ground-truth versus DL derived functional metrics; EDV, ESV, SV, EF, and myocardial mass at ED and ES (n = 25).

A negative correlation was found between blood pool Dice and volume errors, with stronger correlation for EDV than ESV (-0.61 vs -0.45, respectively). Similarly, a negative correlation was also found between myocardial Dice and myocardial mass errors, with stronger correlation for ED than ES (-0.61 vs -0.50, respectively).

Although not necessary for this study, the DL segmentation model can be applied to all timeframes of the SAX cine to enable rapid calculation of time-volume curves, as shown in Figure 4. These curves can be seen to be smooth and can be used to easily extract EDV (maximum of the time-volume curve) and ESV (minimum of the time-volume curve), as is performed in the web-based application.

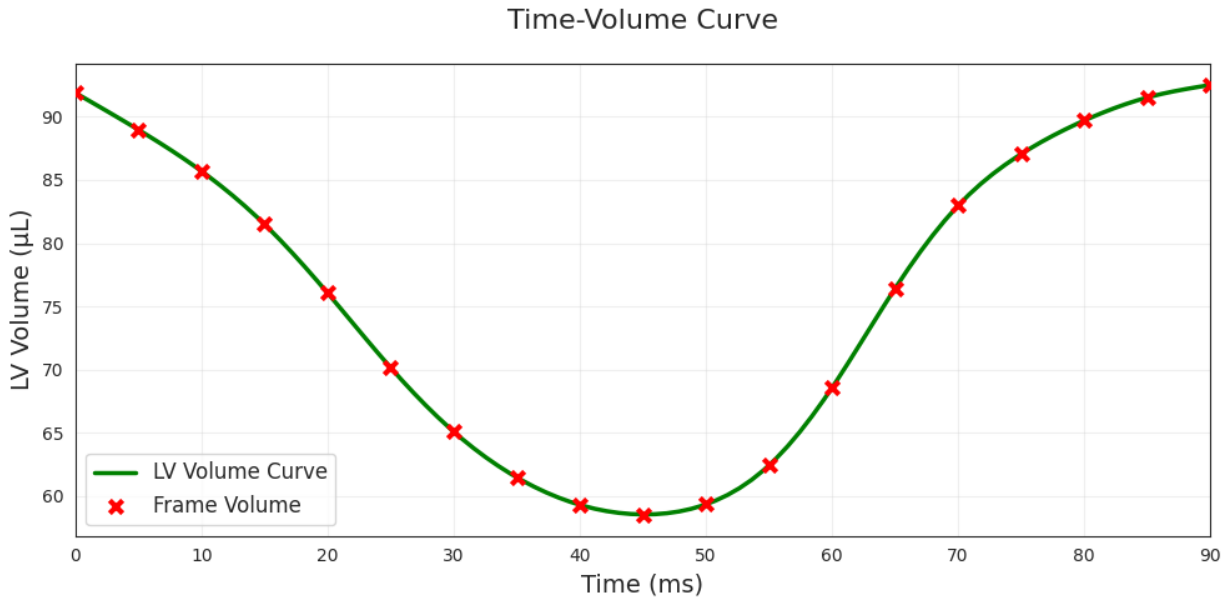

Figure 4: An example LV time-volume curve from one mouse, which can be easily extracted from the DL segmentation model applied to all timeframes and all slices.

Web-Based Inference

The *Hugging Face Spaces* platform enables straightforward upload of pre-clinical cardiac MRI data. For a complete cine short-axis stack (~200 2D images encompassing all slices

and cardiac timeframes) the complete application takes ~3-4 minutes. It accurately identifies ED and ES, and generates left ventricular blood pool and myocardial volumes which are consistent with those obtained from the native Python implementation. In addition, the platform provides visual overlays of the segmentation masks on the cine images, offering users an intuitive means of assessing segmentation quality (Figure 5).

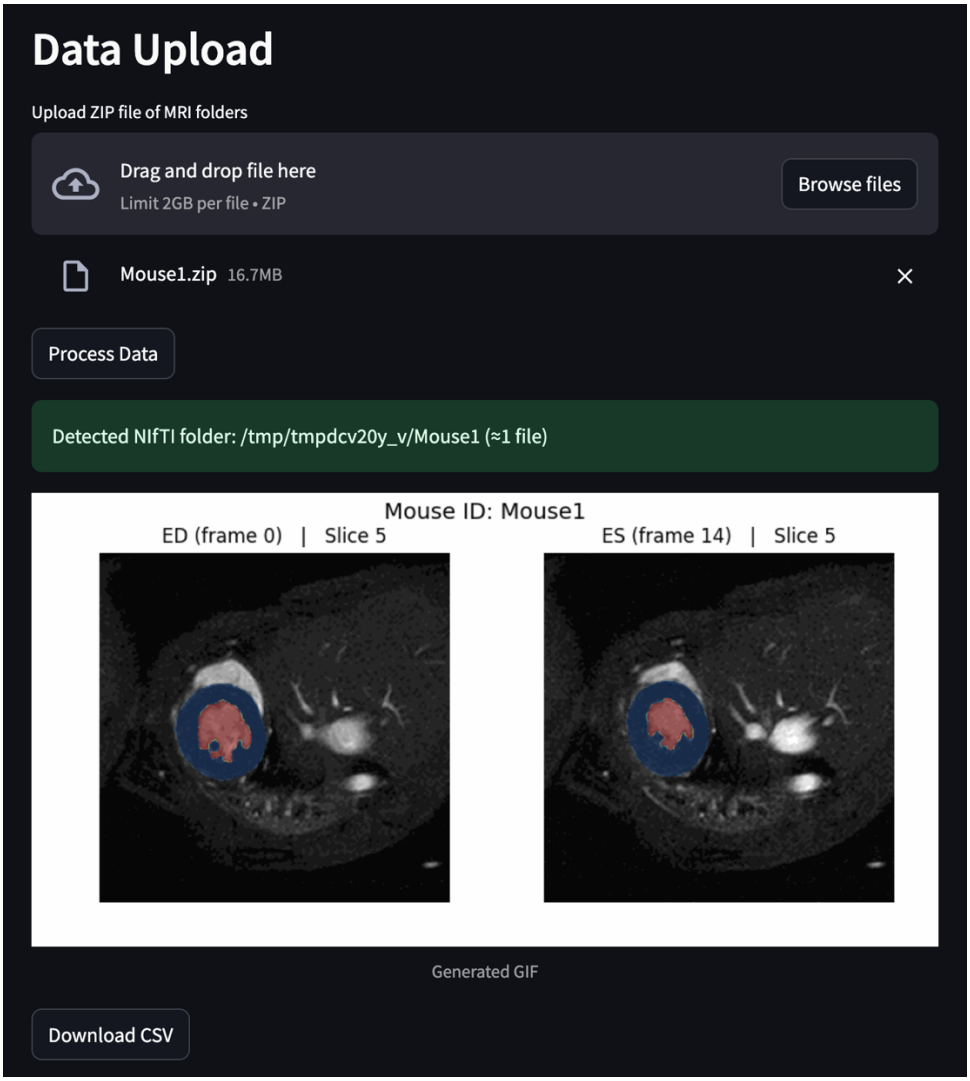

Figure 5: The web-based inference application. The user uploads a zip file containing their NIfTI files, and then selects 'Process Data'. When the application has finished GIFs of the imaging stacks at ED and ES are show with the segmentations overlaid (blood pool in red and myocardium in blue). The application also enables the user to download a .csv file contains the blood pool and myocardial volumes at each time frame.

DISCUSSION

This paper presents the first publicly-available pre-clinical cardiac MRI dataset, as well as the first open-source deep learning model for cardiac segmentation of pre-clinical MRI images. By enabling full access to this state-of-the-art dataset and DL segmentation model, we aim to encourage development of robust segmentation algorithms in mouse MRI. Providing benchmarked segmentations will enable other groups to fairly compare their algorithms, driving the development of innovative pre-clinical cardiac MRI analysis.

Cardiac MRI in mice is particularly important in identifying gene or protein targets for new therapies. Therefore rapid, reliable and reproducible cardiac segmentation is crucial in enabling precise measurement of ventricular volumes and mass, which are essential for assessing treatment efficacy and understanding disease mechanisms. Although manual segmentations are commonly performed in pre-clinical studies, it is slow and has high levels of intra- and inter-observer variability. Instead, DL models offer fast and reproducible segmentations.

Although the use of DL segmentation in human cardiac MRI images is increasing in popularity, it is not possible to apply these models to pre-clinical cardiac MRI images. In fact, a recent paper showed that even between primate groups where the heart size and structures are similar, DL segmentation models do not generalize well to unseen species [24]. Specifically, DL segmentation models trained on cardiac MRI images from rhesus macaques, did not perform well when applied to cardiac MRI images from baboons, demonstrating the need for species-specific segmentation models [24].

Although a few studies have investigated DL segmentation models for mice, they have used private cohorts with few animals (< 21 mice) [13-15]. Instead, our publicly-available dataset contains a significantly larger number of animals (130 mice), with a wide range of myocardial phenotypes. This large, diverse dataset will enable the trained DL

segmentation models to better generalize to unseen data and accurately segment cardiac structures across various pathologies and image qualities.

Open-source DL Segmentation Model

To demonstrate the utility of our publicly-available dataset, we implemented a UNet3+ architecture [20], which has also been made open-source. The model demonstrated good agreement with ground-truth manual segmentation for LV ventricular volumes, ejection fraction, stroke volumes and myocardial mass with results comparable to other DL segmentation models for cardiac MRI segmentation in mice [13-15]. At inference, the DL model was able to segment the LV blood pool and myocardium, on all slices and all timeframes of a SAX cine dataset in < 4.6 s. This enables rapid calculation of smooth LV time-volume curves, with good agreement in ventricular volumes and mass compared to ground-truth segmentations. From these time-volume curves, other metrics including peak ejection rate and peak filling rate could easily be extracted, which can provide additional insight into ventricular function [25, 26]. Additionally, wall thickness measurements, septal curvature and strain quantification could be extracted from the segmented data to further enhance clinical assessments.

Limitations

The main limitation of this study is that ground-truth manual segmentations were performed by a single operator. This was done to ensure consistency of segmentation but could result in biases in DL models. However, we believe that the benefit of having a single adequately trained observer (with expert review when necessary) is that the DL model is trained on highly homogeneous data. Additionally, our publicly-available dataset is from a single centre which may lead to bias and reduced performance of DL models when applied to data from different scanners. Finally, in this paper only segmentations from the LV are performed.

Conclusion

To conclude, we present the first publicly-available pre-clinical cardiac MRI dataset, as well as the first open-source deep learning model for cardiac segmentation in mice. In addition, we have provided an open-source, simple-to-use, free-to-use web-based application to perform inference using our model. We believe that these tools are instrumental in driving advancements and innovations in pre-clinical cardiac MRI, by enabling independent validation and benchmarking of models against each other. We hope this will lead to widespread integration of DL segmentation tools in pre-clinical cardiac MRI, increasing reproducibility, speeding-up post-processing and aiding in the development of novel therapies for cardiovascular disease.

ABBREVIATIONS

CLAHE	Contrast-limited adaptive histogram equalization
DL	Deep learning
ECG	Electrocardiogram
ED	End-diastole
EDV	End-diastolic volume
EF	Ejection fraction
ES	End-systole
ESV	End-systolic volume
GRE	Gradient echo
LV	Left ventricular
MRI	Magnetic resonance imaging
SAX	Short-axis
SNR	Signal-to-noise ratio
SV	Stroke volume
UCL	University College London
vCPU	Virtual Central Processing Unit

ACKNOWLEDGEMENTS

This work was supported by a British Heart Foundation PhD studentship (FS/4yPhD/F/20/34134) and a BHF Accelerator award (AA/18/6/34223). JAS is funded by a UK Research and Innovation (UKRI) Future Leaders Fellowship (grant MR/S032290/1). DJS. is funded by BHF FS/15/33/31608, FS/SBSRF/21/31020, RM/17/1/33377. RC is supported by a UCL UKRI Centre for Doctoral Training in AI-enabled Healthcare studentship (EP/S021612/1). WNH is funded by the Public Service Department of Malaysia.

REFERENCES

1. Heijman E, et al. Evaluation of Manual and Automatic Segmentation of the Mouse Heart from Cine Mr Images. J Magn Reson Imag. 2008;27(1):86.
2. Schneider JE, et al. Fast, High-Resolution in Vivo Cine Magnetic Resonance Imaging in Normal and Failing Mouse Hearts on a Vertical 11.7 T System. J Magn Reson Imag. 2003;18(6):691.
3. Luijnenburg SE, et al. Intra-Observer and Interobserver Variability of Biventricular Function, Volumes and Mass in Patients with Congenital Heart Disease Measured by Cmr Imaging. The International Journal of Cardiovascular Imaging. 2010;26(1):57.
4. Chen C, et al. Deep Learning for Cardiac Image Segmentation: A Review. Front Cardiovasc Med. 2020;Volume 7 - 2020.
5. Bhan A, et al. An Assessment of Machine Learning Algorithms in Diagnosing Cardiovascular Disease from Right Ventricle Segmentation of Cardiac Magnetic Resonance Images. Healthcare Analytics. 2023;3:100162.
6. Bernard O, et al. Deep Learning Techniques for Automatic Mri Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Trans Med Imag. 2018;37(11):2514.
7. Tao Q, et al. Deep Learning–Based Method for Fully Automatic Quantification of Left Ventricle Function from Cine Mr Images: A Multivendor, Multicenter Study. Radiology. 2019;290(1):81.
8. Attar R, et al. Quantitative Cmr Population Imaging on 20,000 Subjects of the Uk Biobank Imaging Study: Lv/Rv Quantification Pipeline and Its Evaluation. Med Image Anal. 2019;56:26.
9. Fernández-Llaneza D, et al. Towards Fully Automated Segmentation of Rat Cardiac Mri by Leveraging Deep Learning Frameworks. Sci Rep. 2022;12(1):9193.
10. Gondova A, et al. Convolutional Neural Networks for Automated Cmr Image Segmentation in Rats with Myocardial Infarcts. bioRxiv. 2020:2020.12.01.405969.

11. Niglas M, et al. Automated Bi-Ventricular Segmentation and Regional Cardiac Wall Motion Analysis for Rat Models of Pulmonary Hypertension. Pulm Circ. 2025;15(2):e70092.
12. Gilson WD, et al. Cardiac Magnetic Resonance Imaging in Small Rodents Using Clinical 1.5 T and 3.0 T Scanners. Methods. 2007;43(1):35.
13. Xu H, et al. Fully Automated Segmentation of the Left Ventricle in Small Animal Cardiac Mri. MIDL Preprint. 2018.
14. Hammouda K, et al. A New Framework for Performing Cardiac Strain Analysis from Cine Mri Imaging in Mice. Sci Rep. 2020;10(1):7725.
15. Zufiria B, et al., editors. Fully Automatic Cardiac Segmentation and Quantification for Pulmonary Hypertension Analysis Using Mice Cine Mr Images. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021 13-16 April 2021.
16. Jasmin NH, et al. Myocardial Viability Imaging Using Manganese-Enhanced Mri in the First Hours after Myocardial Infarction. Advanced Science. 2021;8(11):2003987.
17. Ronneberger O, et al., editors. U-Net: Convolutional Networks for Biomedical Image Segmentation2015; Cham: Springer International Publishing.
18. Isensee F, et al. Nnu-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. Nature Methods. 2021;18(2):203.
19. Chen J, et al. Transunet: Rethinking the U-Net Architecture Design for Medical Image Segmentation through the Lens of Transformers. Med Image Anal. 2024;97:103280.
20. Huang H, et al., editors. Unet 3+: A Full-Scale Connected Unet for Medical Image Segmentation. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2020: IEEE.
21. Yao T, et al. A Deep Learning Pipeline for Assessing Ventricular Volumes from a Cardiac Mri Registry of Patients with Single Ventricle Physiology. Radiol Artif Intell. 2024;6(1):e230132.
22. Buslaev A, et al. Albumentations: Fast and Flexible Image Augmentations. Information. 2020;11(2):125.

23. Wolf T, et al. Huggingface's Transformers: State-of-the-Art Natural Language Processing. arXiv preprint arXiv:191003771. 2019.
24. Ramedani M, et al. Deep Learning-Based Automated Segmentation of Cardiac Real-Time Mri in Non-Human Primates. Computers in Biology and Medicine. 2025;189:109894.
25. Soldo SJ, et al. Mri-Derived Ventricular Volume Curves for the Assessment of Left Ventricular Function. Magn Reson Imag. 1994;12(5):711.
26. Stuckey DJ, et al. Novel Mri Method to Detect Altered Left Ventricular Ejection and Filling Patterns in Rodent Models of Disease. Magn Reson Medicine. 2008;60(3):582.