**Automatic Extraction of Pathologic Biomarkers from Atrial Magnetic Resonance Imaging using Convolutional Neural Networks**

**Author Information**

Zhaohan Xiong1, Aaqel Nalar1, James Kennelly1, Martin K. Stiles, MBChB, PhD2, Steffen Peterson, MD, DPHIL3, Jichao Zhao, PhD1

### Affiliations

**1Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand**

**2Waikato Clinical School, University of Auckland, Auckland, New Zealand**

**3Center for Advanced Cardiovascular Imaging, NIHR Barts Biomedical Research Center, Queen Mary University of London, London, United Kingdom**

### Contributions

Z.X prepared the data, performed data analysis, and drafted the manuscript. A.N and J.K prepared the data and edited the manuscript. M.K.S and S.P provided the data used in this study. J.Z reviewed and finalized the manuscript.

### Corresponding author

Correspondence to: Jichao Zhao (e-mail: j.zhao@auckland.ac.nz, Private Bag 92019, Auckland 1142, New Zealand).

**Abstract**

Accurate quantification of cardiac phenotypes is a crucial step in the diagnosis and treatment of cardiovascular diseases. Current clinical methods are expensive and labour-intensive, relying on multiple imaging modalities and invasive procedures to estimate phenotypes. Here, we present a fully automated framework named AtriaNet which computes the anatomical structure, wall thickness variation, and fibrosis distribution of human atria directly from MRIs. AtriaNet was validated on the world’s largest cardiac cine-MRI (UK Biobank) and contrast-enhanced MRI datasets, surpassing prior state-of-the-art approaches. Further testing on an independent clinical dataset revealed the AtriaNet predictions were accurate when evaluated against clinical methods, and had significantly greater efficiency. The increased efficiency and accurate phenotyping may lead to improved clinical guidance of patients with cardiovascular diseases. As AtriaNet is the first approach of its kind capable of computing multiple phenotypes for any cardiac imaging modality, our robust framework may be transferrable to other challenging medical segmentation tasks.

**Introduction**

Cardiovascular diseases (CVDs) are the largest contributors to the global death toll. An estimated 17.9 million people die from CVD each year, which accounts for 31% of all deaths worldwide1. Recent technological advances in non-invasive medical imaging have led to significant improvement in medical diagnosis, patient stratification, and clinical treatment of CVDs2-4. Among these, cardiovascular magnetic resonance imaging (MRI) is one of the most widely used technologies for assessing cardiac structures due to its ability to acquire images with distinct contrast between soft tissues, high spatial resolution, and the absence of harmful ionizing radiation5. Quantitative phenotypes derived from MRI, such as the chamber volume6, wall thickness/tissue mass7, cardiac function8 and fibrosis (scar tissue)9, are important biomarkers for the determination of pathological states in CVDs and are crucial for decision making in clinics. For example, the left ventricular (LV) ejection fraction is an important clinical biomarker for the early detection of heart failure and cardiomyopathy10. In addition, the left atrial (LA) volume is a key indicator for the severity of cardiac arrhythmias6, such as atrial fibrillation (AF). In spite of this, extraction of patient-specific phenotypes from MRIs remains a complex task requiring comprehensive expertise and intensive manual labour3,4,11. This results in important analyses often being impractical and inaccessible during patient assessments. This has been the major limiting factor for both the treatment of CVDs and research into large-scale MRI cohorts to investigate imaging phenotypes at a population level robustly.

Deep learning, particularly convolutional neural networks (CNNs), has recently become the most popular choice for image analysis, including automatic segmentation of MRIs to quantify phenotypes12. Many benchmarking studies have been conducted in recent years investigating the optimal deep learning algorithms for analyzing cardiac structures. A 2017 study for the segmentation of LV and right ventricle (RV) from 150 3D MRIs showed that CNNs significantly outperformed traditional atlas-based and shape-model methods13. By analyzing algorithms from 10 teams, the study revealed that the popular U-Net CNN architecture which was initially designed for histological image segmentation was the most effective14. In a separate whole heart segmentation study15, the results of 12 teams also showed the superiority of CNNs on 60 3D MRI and computed tomography data, which were shown to achieve more stable and higher scores across all chambers of the heart. More recent studies have focused on the use of contrast-enhanced MRIs (CE-MRIs) due to their clinical utility for visualizing diseased cardiac tissue, although their analysis is more challenging due to the attenuated contrast in nondiseased tissue. In 2018, a study for LA segmentation16 from 154 3D CE-MRIs showed the best teams designed U-Net CNNs with additional residual connections and custom optimization loss functions to surpass the vanilla U-Net architecture. Similar findings were obtained by a 2019 benchmarking study17 for LV and RV segmentation from 45 CE-MRIs, where the highest-scoring team utilized an enhanced U-Net approach.

Despite these global efforts, the aforementioned studies have mostly focused on performing cardiac segmentation on small in-house datasets of 100 to 200 independent samples. Recent studies have also only investigated the segmentation of the heart specifically to derive chamber volumes to measure cardiac function18. No studies have developed a robust pipeline for extracting all key cardiac imaging phenotypes such as cardiac anatomy, chamber diameter, volume, wall thickness, and fibrosis. More importantly, no studies have demonstrated the robustness and generality of their approach on multiple independent datasets, which is essential for utilizing it in clinical settings. The UK Biobank is a prospective cohort study consisting of 500,000 participants designed specifically for studying the risk factors of common diseases including CVDs19. There is an onging imaging substudy of the UK Biobank using MRI in a target sample size of 100,000. Currently, cine-MRIs from 5,000 individual patients have been manually annotated to derive the phenotypes of the cardiac structure including the ventricle, atria and aorta, leading to the largest cardiac cine-MRI dataset in the world. In the past 20 years, a clinical group at The University of Utah has been leading a global initiative to phenotype cardiac fibrosis by utilizing CE-MRIs of the heart9, resulting in the development of the world’s large cardiac CE-MRI dataset with over 13,000 manually annotated 2D LA. This dataset in particular is of exceptionally high quality due to the careful annotations and high-resolution imaging used, allowing a clear representation of the entire heart and its key structures including diseased tissue. While the LA has been heavily investigated, no established studies exist for phenotyping the RA. Our team has manually annotated the RA in the CE-MRI dataset20. The availability of these two large and high-quality datasets presents an ideal setting for extensive development and validation of more robust CNNs.

In this study, as a first attempt to address the essential issues in the field, we have developed a fully automatic framework, with the alias *AtriaNet*, to analyze both atrial chambers (LA and RA) of the heart. More importantly, the computational pipeline is the first stand-alone framework capable of providing cardiac quantitative phenotyping including the anatomical structure, diameter, volume, wall thickness variation, and fibrosis distribution for any cardiac imaging. AtriaNet was developed and validated on the world’s largest cine-MRI and CE-MRI datasets, UK Biobank and University of Utah, and has the potential to be extended to other cardiac structures in a wider context. More importantly, we further validated the proposed method for its clinical efficacy in guiding the treatment of AF, the most common cardiac arrhythmia. We were the first to automatically phenotype both atrial chambers from CE-MRIs as AF is a bi-atrial disease and prior studies are focused on the LA only. Our framework was tested on an independent clinical MRI dataset at Waikato Hospital to demonstrate its applicability on cross-centre imaging21. We also validated the accuracy of AtriaNet in estimating clinical biomarkers by comparing the predictions to patient biomarkers recorded during clinical practices alongside the MRI acquisition.

When compared with other approaches in the field, the AtriaNet framework demonstrates multiple advantages and improvements. The designed CNN architecture is capable of estimating multiple cardiac phenotypes using the same network architecture and can be applied to both 2D and 3D imaging. Our method is also the first method for bi-atrial chamber analysis. It achieves high accuracy and can obtain similar accuracy levels even with lower sample size datasets. In terms of computation time, it is orders of magnitude faster than current clinical approaches. Lastly, we demonstrate the performance consistency of AtriaNet when applied to cross-centre imaging, and its effective transferability for computing accurate quantitative phenotypes from independently acquired clinical datasets. Overall, this study is a very important step towards improved and more efficient patient-specific diagnosis and treatment, and the developed methods may potentially be transferable to other imaging tasks, and lead to advancements in cardiac imaging phenotyping for the treatment of other CVDs.

**Results**

### AtriaNet

The AtriaNet pipeline consisted of three stages: pre-processing, deep learning, and post-processing (Fig. 1a). The pre-processing stage was used to normalize the image intensities and slice image volume into 2D images for 3D data. The deep learning stage consisted of two CNNs, both using the same architecture (Fig. 1b). The first CNN localized and cropped a focused sub-region containing the atria to alleviate class imbalance, reduce task complexity and computational cost. The second CNN then performed targeted prediction on the sub-region to produce the desired output based on the given input labels. Extensive hyper-parameter tuning on the number of layers, layer depth, and kernel size of the CNN architecture (Supplementary XXX) was performed to optimize the performance. The post-processing stage involved padding the cropped outputs to the original input dimensions, and in the cases where the input was 3D, the individual slices were stacked and reconstructed to the original 3D dimension.

The output of the second CNN was adjusted to predict different cardiac biomarkers for image phenotyping. For the atrial anatomy, RA/LA diameter, volume and fibrosis, a softmax activation function was imposed at the output layer to restrict values to probabilistic maps. The atrial anatomy was obtained through direct LA and RA segmentations produced by AtriaNet. The RA/LA diameter and volume were then calculated from the atrial cavity segmentations. Atrial fibrosis was also obtained through direct segmentation of the fibrotic and non-fibrotic pixels in the LA and RA walls. To obtain the atrial wall thickness, a rectified linear unit was used at the output layer to enforce positive and continuous values, leading to AtriaNet directly outputting a wall thickness distribution map.

### Multi-Center Data for Validation

Clinical datasets from three centers (UK Biobank19, University of Utah9, and Waikato Hospital21) were used in this study. The UK Biobank dataset contained 4,860 2D long-axis cine-MRIs with two-chamber and four-chamber perspectives and a spatial resolution of 1.8 × 1.8 mm2. Each sample was manually annotated with the LA and RA cavities at end-diastole and end-systole. The Utah dataset contained 13,552 2D CE-MRIs with a spatial resolution of 0.625 × 0.625 mm2, manually annotated with cavities and walls of the LA and RA. The Waikato dataset contained 968 2D CE-MRIs, with the same labels and spatial resolution as Utah’s. 2D echocardiograms and 3D low-voltage maps obtained via a commercial electro-anatomical mapping system (CARTO, Biosense Webster, United States) were used to validate the atrial diameter, volume, and fibrosis predictions from AtriaNet. Numerical solutions of the atrial wall thicknesses were used to validate the wall thickness estimations from AtriaNet.

The three datasets were individually split into training, validation, and testing sets. The UK Biobank cine-MRIs were split into 82% training, 6% validation, and 12% testing. The Utah CE-MRIs were split into 60% training, 10% validation, and 30% testing, and lastly, the Waikato CE-MRIs were only split into 36% training and 64% testing.

### Accurate Atrial Chamber Segmentation

The segmentation accuracy of AtriaNet was evaluated on the three datasets described (Fig. 2). We first quantified the overlap between the human expert annotations and predicted atrial segmentation using the Dice score. We then assessed the anatomical accuracy of the predictions by analyzing surface-to-surface distance (STSD) errors from the ground truth. All results presented were obtained by evaluating AtriaNet on the testing set during development and parameter tuning.

The Dice score of AtriaNet on the 2D UK Biobank cine-MRIs was 95.6% and 96.1% for the LA and RA cavities respectively (Fig. 2a). Expectedly, the Dice scores were 1% to 2% higher for the end-diastole images due to the increased atrial areas in the scans. The low standard deviation of 2% suggested extremely consistent performance for both chambers. The results surpassed prior studies that used the same ratios of training, testing and validation with their datasets22. Furthermore, visualizations showed excellent predictions from AtriaNet when compared with the ground truth, as reinforced by the low mean STSD of 1 mm, or approximately half a pixel (Fig. 2c). In particular, challenging regions such as the atrioventricular valve and pulmonary vein antrum which frequently lack clear anatomical boundaries in MRIs were segmented accurately by AtriaNet.

Segmentations of the atrial chamber in 3D CE-MRIs from Utah and Waikato were more challenging due to the increased anatomical detail present in the high contrast imaging. Despite this increased difficulty, AtriaNet achieved Dice scores of 91.9% and 91.2% for the LA and RA cavities with interquartile ranges between 90% and 93%, showing high precision 3D segmentation across the test datasets (Fig. 2b). Visualizations in 3D further showed the majority of the predicted atrial anatomy was less than 1 pixel (approximately 0.4 mm), from the ground truth (Fig. 2d). The pulmonary veins, which were not labelled in the cine-MRI due to low resolution, contained the highest distance errors in the CE-MRIs. However, the errors were mainly attributed to the predicted veins extended out from the LA chamber at different lengths than the ground truth, a metric which was difficult to define even during the expert annotation. Further 2D comparative visualization revealed other important structures such as the left atrial appendage and atrioventricular valves were also segmented with high precision (Fig. 2e).

The atrial cavities were segmented to enable the derivation of the diameter and volume measurements. However, analysis of atrial wall thickness and fibrosis requires the segmentation of the atrial wall tissue. To extract LA and RA walls from the CE-MRIs, AtriaNet first segmented the epicardial surface to delineate the enclosed tissue mass consisting of both the atrial cavity and wall. The wall was therefore the non-overlap region between the epicardial surface and cavity segmentations. AtriaNet segmented the enclosed atrial cavity and wall with a particularly impressive Dice score of 93%. These observed performances were consistent across both CE-MRI datasets.

Overall, the consistent performance across the three multi-centre datasets demonstrated the generalizability of AtriaNet. In particular, accurate predictions were maintained regardless of using the larger UK Biobank dataset or the smaller Waikato dataset.

### AtriaNet Calculates the Chamber Volume and Diameter with Clinical-Level Precision

### AtriaNet is able to efficiently generate accurate atrial measure calculations directly from its automatic segmentations. We compared the percentage accuracy between the AtriaNet atrial diameter and volume predictions against those calculated from the expert segmentations in all three datasets. In the cine-MRI dataset, the diameter and volume errors were consistent for each atrial chamber, yielding accuracies of 97% for the RA and 96% for the LA (Fig. 3a). The errors in the CE-MRI datasets were marginally lower for the RA at 93% for both measures, and higher for the LA which had a volume accuracy of 94% and a diameter accuracy of 96% (Fig. 3b). Similar to the segmentation evaluation, the accuracies on the CE-MRIs were lower on average, although this was less noticeable, especially for the LA. Surprisingly, AtriaNet achieved higher accuracies for estimating the diameter and volume of the predicted atrial segmentation compared to the accuracy in predicting the actual segmentation themselves. This implied a lower propensity for errors when AtriaNet is used to compute important metrics for phenotyping.

To further validate the clinical utility of our pipeline, we compared AtriaNet chamber measurement predictions with clinical measures reported by radiologists in 2D echocardiography of the same patients at Waikato (Fig. 3c). Specifically, we compared the correlation between the estimated areas in the AtriaNet 3D segmentation predictions against the areas measured from the echocardiography scans. The areas were estimated from the AtriaNet segmentations by reversing the commonly used bi-plane equation given the volume and diameter of the atrial chambers. Fig. 3c shows the relationship between the CE-MRI derived areas and the echocardiography-derived areas. The measurements for the both the LA and RA had statistically significant and strong Pearson’s correlations of 0.8 (p = 0.01). We also compared the correlations directly between the predicted volume/diameter with the echocardiography areas (Supplemental Fig. XXX). Both the volume and diameter were correlated with the area for the LA (correlation > 0.7, p < 0.05). The RA volume also had a high correlation (correlation = 0.8, p = 0.01), while the diameter was weakly correlated (correlation = 0.5, p = 0.19). The strong correlations demonstrated the AtriaNet predictions closely matched the values measured clinically, with slight variations being attributed to the different imaging modalities used.

Overall, the results demonstrated AtriaNet was able to accurately phenotype the atrial volume and diameter. AtriaNet provides a significant improvement in fully automating atrial chamber measurement calculations. Furthermore, as current clinical methods involve estimating the atrial volume with the bi-plane method on 2D echocardiography scans, AtriaNet provides a more robust alternative by analyzing 3D imaging which contains more complete anatomical information.

### AtriaNet Estimates Accurate 3D Wall Thicknesses

We extended AtriaNet to compute 3D atrial wall thickness estimations from the CE-MRI segmentations. The wall thickness variation, particularly in the RA, has been demonstrated to have an important role in arrhythmogenesis as it influences the electrical propagation patterns during episodes of arrhythmia23. Wall thickness also plays an important role in aiding the selection of regions to target during ablation surgery for treating arrhythmia7,24. The CE-MRI dataset from Waikato was manually annotated to account for thickness variation in the atrial wall tissue. The dataset is the first of its kind to contain segmentations of both atrial chambers that accounted for the atrial wall thickness variations in CE-MRIs. Ground truth maps of the 3D wall thickness were obtained by solving computationally expensive differential equations. The mean pixel-wise error was used to evaluate the predicted 3D wall thickness map.

AtriaNet obtained mean pixel-wise errors of 0.4 mm and 0.6 mm for the LA and RA, respectively, which corresponds to the predictions being within one pixel from the ground truth on average. The mean atrial wall thickness value was also estimated with a high accuracy of 97% for both atrial chambers (Fig. 4a). AtriaNet successfully reproduced the overall wall thickness distribution in all hearts. Visualizations of the 3D predictions showed that key anatomical structures, such as the crista terminalis and the pectinate muscles, were accurately captured and also demonstrated the overall trend of the LA being thinner on average relative to the RA, with the LA having less fluctuations in wall thickness throughout the chamber (Fig. 4c). The accuracy of the wall thickness estimation can further be seen in enlarged 2D visualizations (Fig. 4d). Analysis of the wall thickness distribution graphs showed the distribution of the predicted values were slightly narrower than the ground truth. This revealed the pipeline had difficulty capturing extreme thickness values on both tails of the distribution, although this had a minimal impact on the overall accuracy because the amount of tissue at these extremities was low.

The motivation behind designing AtriaNet to predict the wall thicknesses was the extremely labor-intensive pre-processing required and the high computational costs associated with traditional mathematical methods. The generation of the ground truths required the manual delineation of the atrial endocardial and epicardial surfaces for setting up the boundary conditions for the Laplace equations. This process alone can take weeks even by a skilled expert. This, combined with the intensive computation time required to solve the equations, resulted in the entire wall thickness ground truth taking up to two weeks to produce one sample. On the other hand, the complete wall thickness map estimation for one 3D CE-MRI dataset took AtriaNet approximately 10 seconds on average, presenting an over 10000-fold increase in processing speed (Fig. 4b). The high accuracy of our approach and the substantial reduction in processing time can potentially accelerate patient-specific structural analysis in clinics and lead to more efficient and unbiased diagnoses.

### AtriaNet Fibrosis Quantification and Validation with Clinical Electro-Anatomical Mapping

We re-trained AtriaNet to segment the fibrotic and non-fibrotic pixels in the atrial walls of the CE-MRIs from Utah and Waikato. Ground truth atrial fibrosis was obtained using a semi-automatic, adaptive thresholding method to isolate high-intensity pixels, deemed as fibrotic tissue. The accuracy of AtriaNet was evaluated by comparing the predicted fibrosis distribution with the ground truth distribution using Kolmogorov-Smirnov correlations25. Fibrosis segmentations enabled the calculation of the percentage of fibrotic tissue in the wall, an important biomarker which has been shown to predict the success of clinical treatment outcomes. Therefore, we also compared the error in fibrosis percentage calculated from the predictions and ground truths.

AtriaNet obtained correlations of 0.90 and 0.92 (p < 0.05) for the predicted LA and RA fibrosis segmentations respectively in the Utah dataset (Fig. 5a), and a 3.0% and 3.7% error for the predicted fibrosis percentage in the LA and RA respectively. The performance on the Waikato dataset was slightly lower with correlations of 0.88 and 0.91 (p < 0.05) for the LA and RA fibrosis segmentations, while the fibrosis percentages were estimated with an error of 3.9% for the LA, and a highly accurate 1.9% error for the RA. The average fibrosis percentages were fairly consistent in both datasets, with Utah having 18% for both chambers and Waikato having 20.4% for the LA and 19.5% for the RA. Expectedly, the CE-MRIs scans acquired in patients who underwent ablation treatment had on average 3% more fibrosis due to the scarring incurred as a result of the procedure. Visualizations of the AtriaNet fibrosis segmentations also showed excellent matches with the ground truth, and the spatial distribution of fibrotic pixels were accurately reproduced (Fig. 5b). Notably, the regions with a higher fibrosis occurrence were effectively captured, such as the pulmonary vein antrum and mitral valve in the LA, and the superior vena cava in the RA. The only cases with higher errors were ones with a low fibrosis percentage due to the smaller number of fibrotic pixels being more difficult to detect. However, in these instances, the shape of fibrosis distribution was still accurately predicted.

The accuracy of manually assessing the LA fibrosis from in-vivo imaging has often been scrutinized in previous literature26. Varying methods used to define fibrosis results in widely subjective ground truth labels27, compromising further attempts of developing automated methods to predict fibrosis28. We attempted to overcome this issue by comparing our AtriaNet predicted fibrosis maps against low-voltage maps recorded invasively during 3D electro-anatomical mapping of patients at Waikato (Fig. 5c). The low electrical conduction of fibrotic tissue leads to low voltage values being recorded, which could be used to visually assess the accuracy of the fibrosis map generated from the CE-MRIs. Fig. 5c shows three patient data containing the paired CE-MRI and voltage maps. While direct quantitative comparisons were infeasible due to the differing methods used to anatomically reconstruct the LA29, the overall pattern of the predicted fibrosis closely matched the low voltage regions (in red) of the voltage maps. This was particularly the case around the pulmonary veins as well as the mitral valve which contains more fibrotic tissue. While the match between the pairs were not exact due to errors associated with the different LA reconstruction techniques, this experiment successfully demonstrated the ability of the AtriaNet pipeline in quantifying fibrosis in both an accurate and realistic manner.

**Discussion**

The study presents AtriaNet, the first fully automated pipeline for extracting the atrial anatomy, chamber measurements, 3D wall thickness, and distribution of fibrosis. In AtriaNet, we designed a simple, effective CNN-based framework containing two sequentially arranged CNNs for detecting and subsequent phenotyping of the atria. We enhanced the CNNs from its baseline U-Net architecture with residual blocks at every receptive level and pre-activation normalization at every layer. The proposed method was, to the best of our knowledge, validated on the world’s largest cardiac MRI datasets consisting of the UK Biobank 2D cine-MRIs and the University of Utah 3D CE-MRIs. We further tested the efficacy of AtriaNet on our locally acquired clinical CE-MRI dataset and compared it with clinical measures. Overall, AtriaNet achieved excellent LA and RA cavity segmentations on the cine-MRIs, and LA and RA cavity and wall segmentations on the CE-MRIs. Notably, AtriaNet was the first fully automatic approach for RA phenotyping from CE-MRIs. The atrial measurements subsequently calculated from the segmentations yielded accurate atrial diameter and volume estimates, closely correlating with clinically recorded measures in echocardiography reports. We then demonstrated AtriaNet’s ability to accurately compute the bi-atrial wall thickness distribution on the CE-MRIs, with a significant improvement in the efficiency of execution and computational speed over traditional methods. Lastly, AtriaNet was capable of predicting the bi-atrial fibrosis distribution on the CE-MRIs with a high degree of precision, and closely matched low-voltage maps of the atrial fibrosis recorded during non-invasive electro-anatomical mapping.

Segmentation of medical imaging is an important step to delineate the anatomical structures of interest for measuring clinical diagnostic biomarkers. In recent years, extensive research has been conducted for the automatic segmentation of the bi-ventricular chambers’ cavities13,30, which includes the LV wall. Similar to our study, state-of-the-art bi-ventricular segmentation approaches utilize enhanced variants of U-Net18,31, with the use of the two-CNN configuration becoming popular in recent years30,32. In contrast, atrial segmentation has been investigated in significantly fewer studies, mostly focusing on the LA. Prevailing methods involve extensive hyper-parameter tuning of U-Net based architectures to account for the increased difficulty of segmenting the more morphologically complex LA cavity16. In this study, the segmentation capability of our proposed approach is advantageous as AtriaNet is one of few methods able to directly segment the walls of the atrial chambers. Prior studies have mainly focused on cardiac chamber cavity segmentation which is sufficient for calculating chamber measurements33,34. However, the fibrosis distribution and wall thickness variation which are the key factors of dictating patient stratification and treatment planning35,36 can only be extracted from the delineated wall tissue. The atrial walls are significantly more difficult to capture as the walls are approximately two and four times thinner for the RA and LA respectively23 compared to the LV. Currently available methods initially estimate the atrial wall geometry by extrapolating the cavity, leading to inaccurate extraction of the fibrosis37 and wall thickness38. Studies which perform accurate fibrosis quantification are still limited to manual wall segmentations39, while automated cardiac wall thickness estimation has only been attempted for the thick LV wall40. In comparison, our proposed AtriaNet directly produced accurate bi-atrial wall segmentations with our simple architecture designed specifically to detect the pixels along the thin atrial wall, which occupies less than 0.01% of the MRI. Further, our methods has also been extensively validated on the world’s largest cardiac MRI dataset, being more robust than similar studies with small datasets41. The precise atrial wall segmentations subsequently led to accurately predicted wall thickness and fibrosis biomarkers, as seen from both the quantitative and clinical validation results.

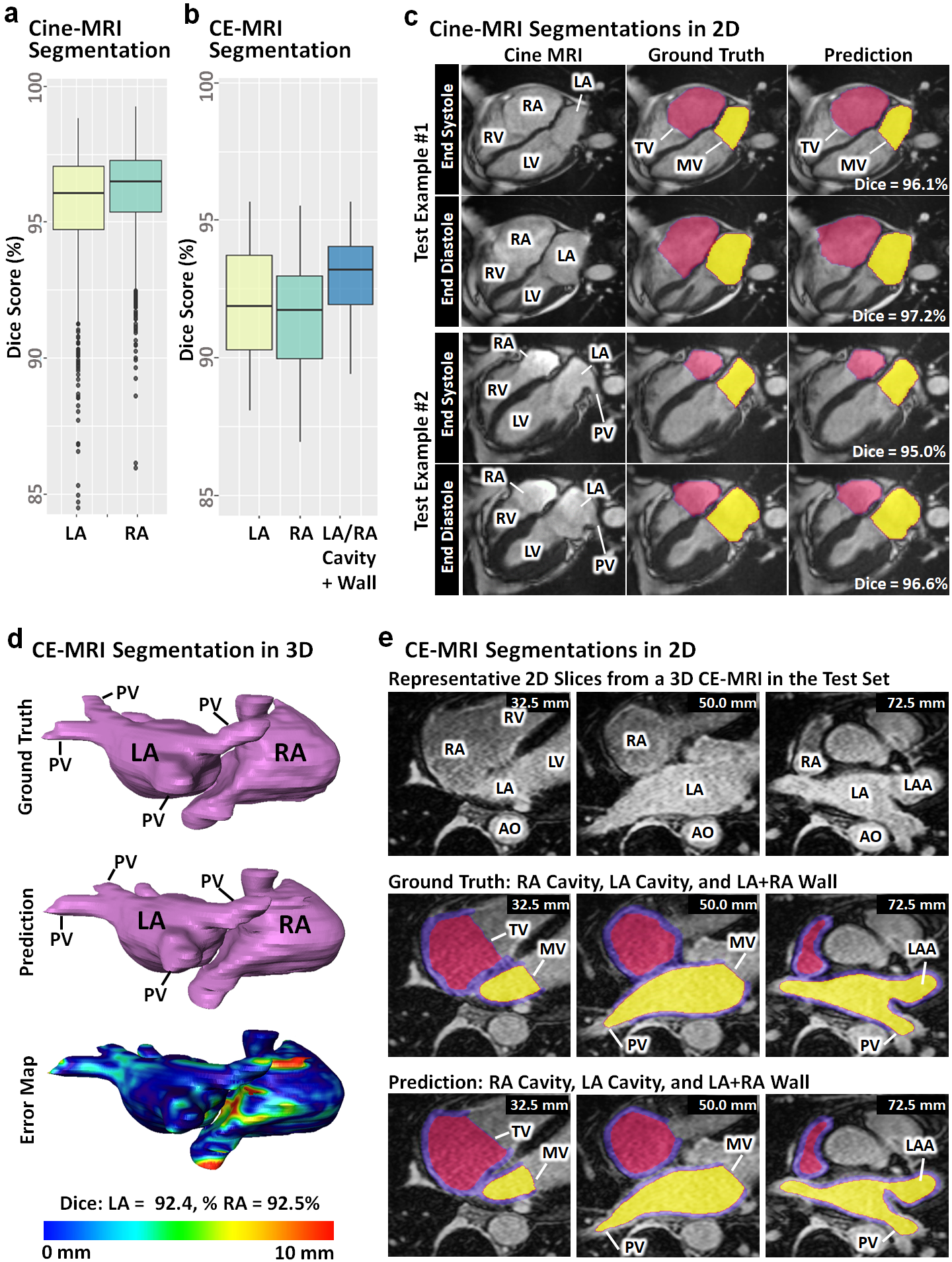
The simple design of AtriaNet allowed the pipeline to perform multiple difficult tasks without significant modification. The activation function of the last layer and the loss function of the second CNN were the only parameters altered when executing the three different phenotyping tasks. Since these were fixed parameters, the trained CNN weights could be easily shared across different tasks and datasets when data availability was limited. The benefits were more pronounced when developing the model for wall thickness estimation on the Waikato CE-MRI dataset, as the CNNs were initialized with pre-trained weights from segmenting the larger Utah dataset to accelerate convergence. The usability of our approach allowed our pipeline to be more efficient by reusing the same model. This also demonstrated the robustness of the proposed configuration of two sequential CNNs for image analysis in general, as our approach was shown to be independent of any specific input, label, or imaging dimension. Computational costs were minimized through the exclusive use of 2D CNNs for both 2D and 3D tasks, as 3D convolutions result in an exponentially greater number of parameters which increases the difficulty of training and convergence42. The ease of convergence of AtriaNet was important as the atrial prediction tasks contained data with a high class-imbalance of over 99.9%, increasing the likelihood of locally optimal parameters of entirely zero weights43. These issues were further mitigated with the combination of weighted loss functions to target positive pixels, pre-trained weights to initialize each training session, and convolutions with a relatively lower depth a minimal number of trainable parameters. The lightweight design also allows AtriaNet to be executed on most standard computer hardware without compromising runtimes.

Future improvements to our study would involve increasing the data quality for validation, especially for the clinical validation of the atrial phenotypes. The clinical data available would also ideally be specifically acquired during the trials for evaluation against the predictions from CE-MRIs. For example, in future studies, the low-voltage maps recorded would be projected onto the CE-MRI segmented anatomy during the clinical recording phase to allow for direct comparisons of the fibrosis maps44. Further studies should also be set up to develop more samples of wall thickness variation. The development of semi-automatic, but accurate methods of wall thickness delineation would also be beneficial as current manual processes are too time consuming to be applied on a large scale. We also aim to acquire clinical information on our MRI datasets to be able to associate the predicted phenotypes retrospectively with patient outcomes and perform survival analysis. Lastly, the generalizability of AtriaNet would be further validated on more diverse imaging modalities such as computed tomography scans or imaging on other cardiac structures.

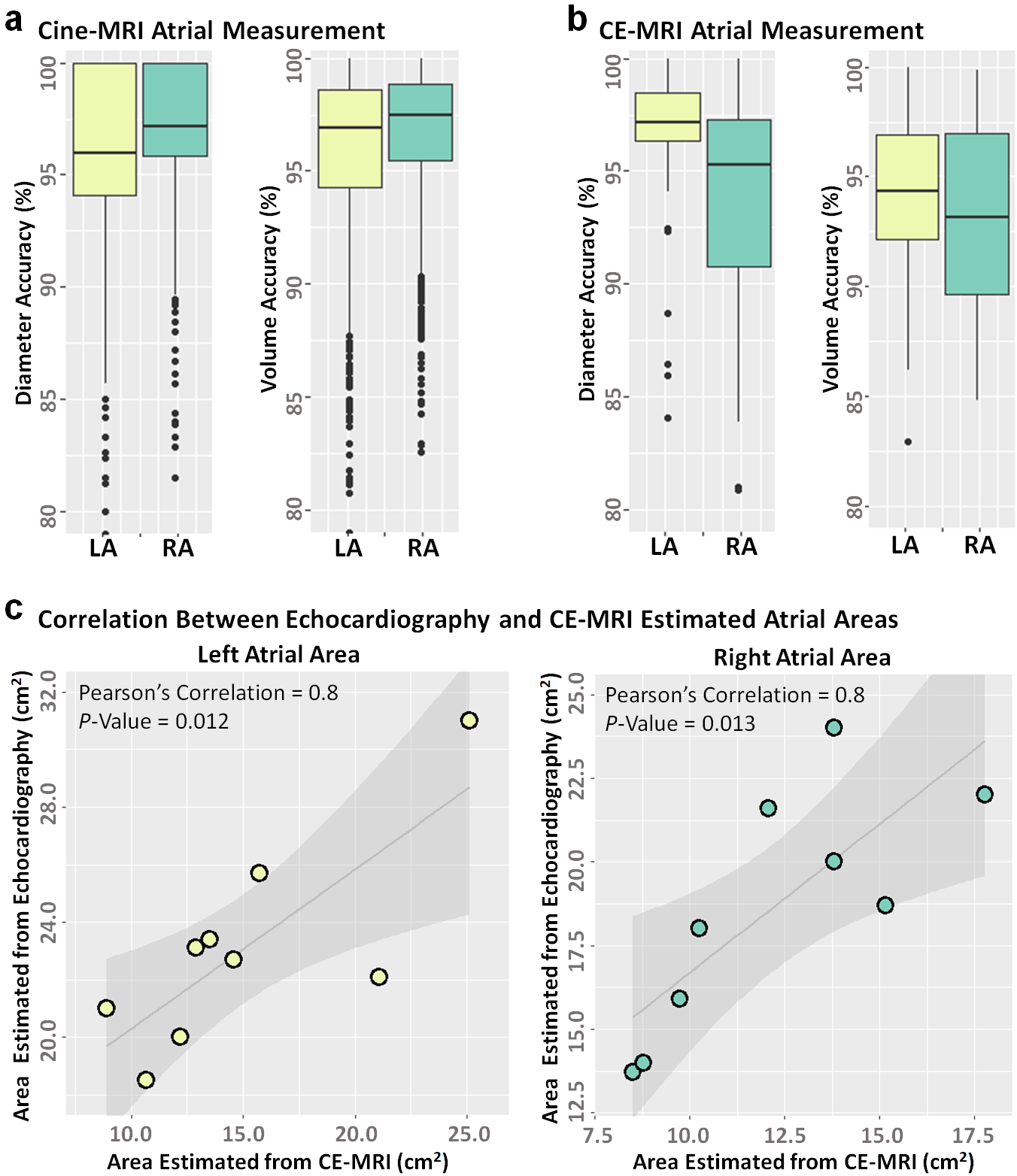
Overall, we have proposed a robust pipeline for automated atrial phenotyping which was validated on diverse and cross-centre imaging datasets. AtriaNet was shown to produce high quality atrial segmentations, wall thickness estimations, and fibrosis quantifications which are crucial prerequisites during clinical assessment and treatment. The increased efficiency and accurate phenotyping may lead to the development of improved clinical diagnosis, patient stratification, and clinical guidance of patients with CVDs. The insights gained from our study may also impact the wider imaging community for transferring our simple yet effective framework to other image phenotyping tasks.

### 

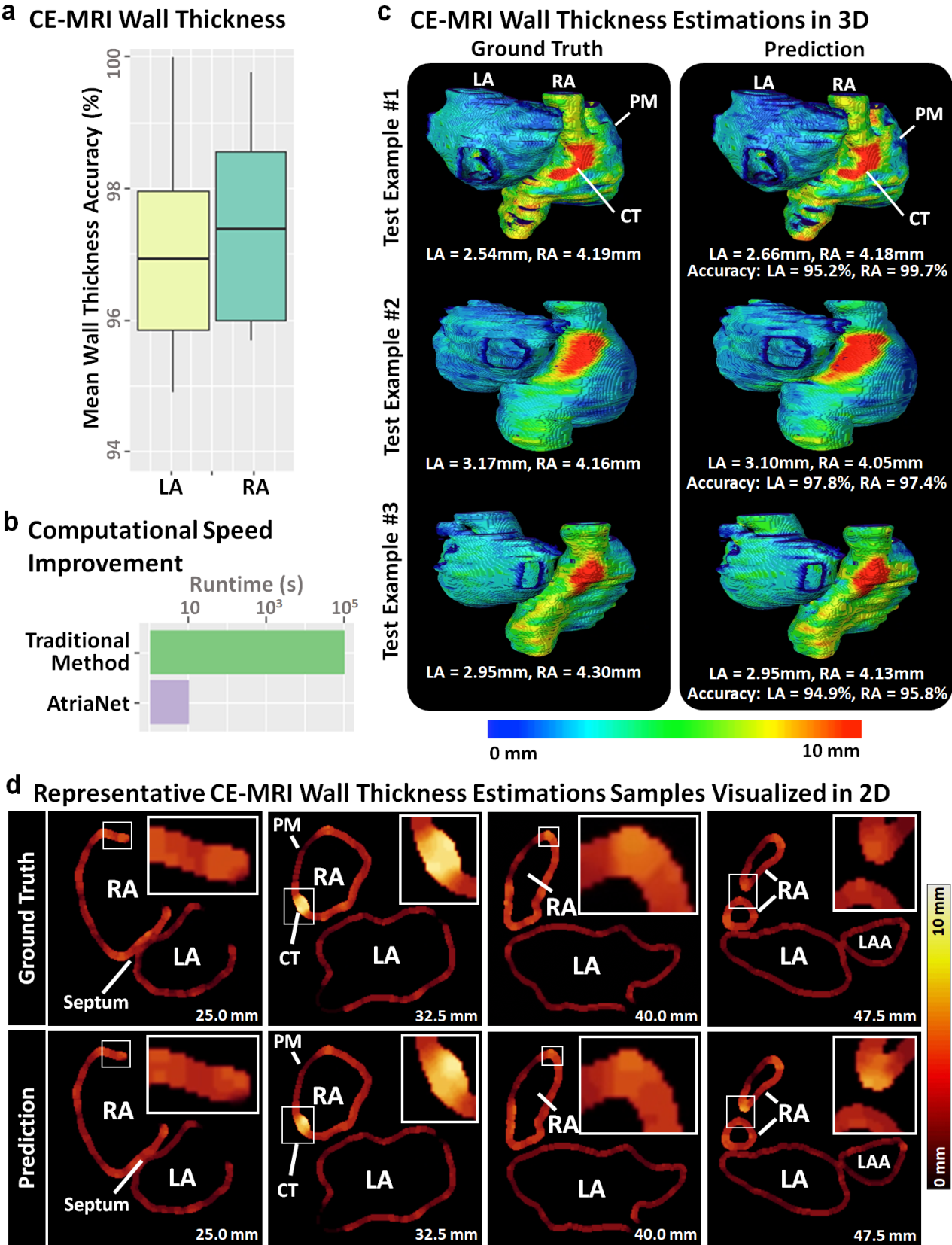
**Fig. 1 | Illustration of the AtriaNet pipeline for automatic extraction of pathologic biomarkers from cardiac magnetic resonance imaging (MRI). a**, The two-stage convolutional network (CNN) configuration of AtriaNet for fully automatic left atrial (LA) and right atrial (RA) segmentation, wall thickness estimation, and fibrosis quantification. The first stage detects the region of interest (ROI) from the MRI containing the atrial chamber. The second stage then performs the analysis on the subsequent ROI to produce the final output. **b**, The architecture of the CNN used both stages of AtriaNet. BN, batch normalization; Conv, convolution; PReLU, parametric rectified linear unit.



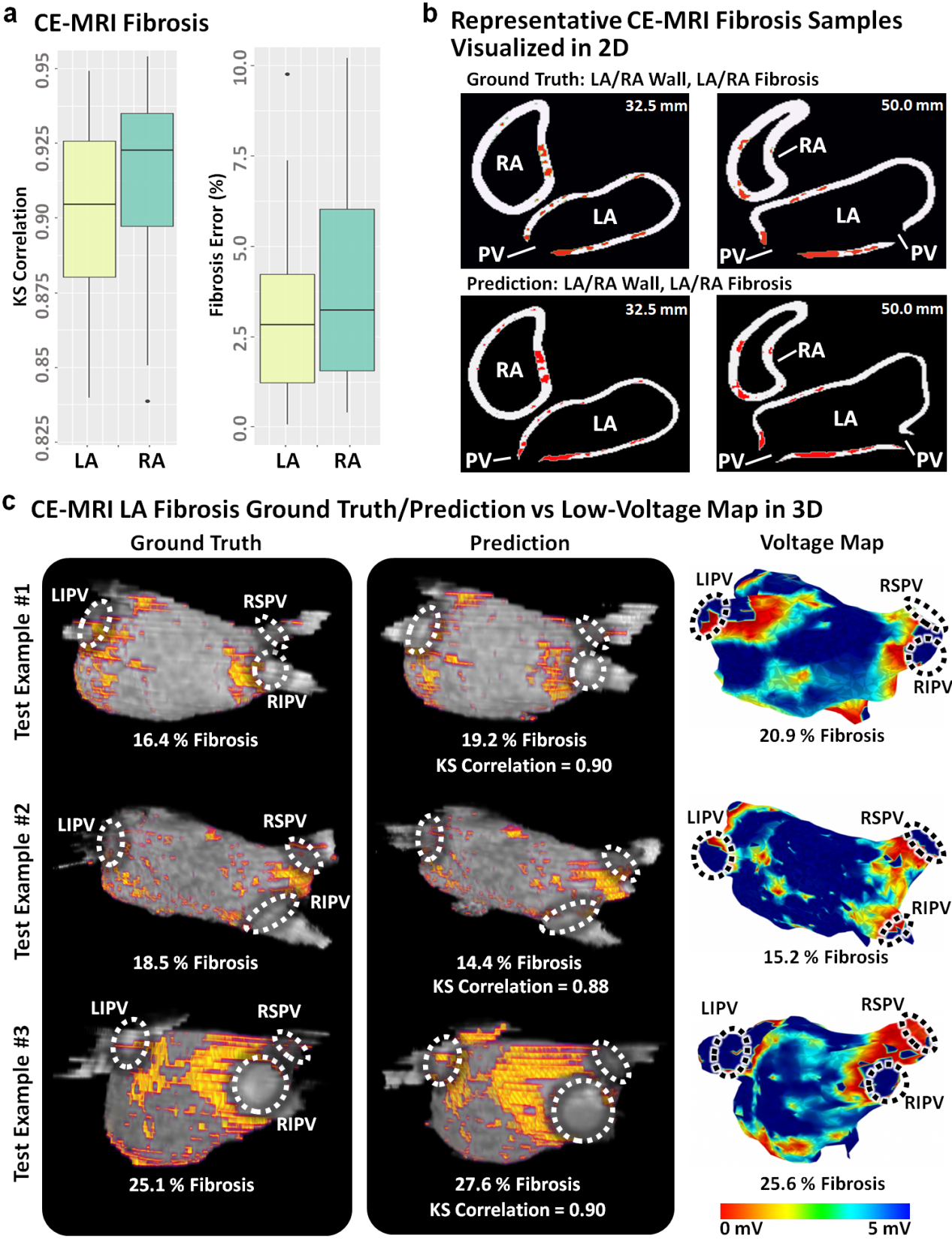
**Fig. 2 | Segmentation performance of AtriaNet on 2D cine-magnetic resonance imaging (MRI) and 3D contrast-enhanced (CE)-MRI.** Quantitative results from the left and right atrial (LA/RA) segmentation of **a**, cine-MRIs and **b**, CE-MRIs measured with the Dice score. **c**, Qualitative results from the segmentation of the two representative cine-MRIs samples. Each row represents the same cine-MRI sample and each column represents the raw image, ground truth, and AtriaNet prediction for comparison. **d**, Qualitative results from segmenting a representative CE-MRI in 3D, comparing the ground truth, prediction, and surface to surface distance error map. **e**, Qualitative results for 2D CE-MRI segmentation for the same patient in (**d**)at representative slices. Each column represents the same slice sample and each row represents the raw CE-MRIs, ground truth, and prediction. The spatial position of each CE-MRI slice is labelled for each sample. AO, aorta; LV, left ventricle; LAA, left atrial appendage; MV, mitral valve; PV, pulmonary vein; RV, right ventricle; TV, tricuspid valve.



**Fig. 3 | The accuracy of AtriaNet for atrial measurement calculation from 2D cine-magnetic resonance imaging (MRI) and 3D contrast-enhanced (CE)-MRI, and clinical validation using clinical echocardiography.** Quantitative results for the percentage accuracy of estimating the left and right atrial (LA/RA) diameter and volume from **a**, cine-MRIs and **b**, CE-MRIs. **c**, Correlation of the LA and RA areas calculated from clinical assessment of echocardiography scans and computed automatically by AtriaNet from the CE-MRIs. The linear regression line is shown with a standard error shadow. The correlation coefficients and statistical significance level is also shown.



**Fig. 4 | Evaluation and visualization of the wall thickness estimation accuracy of AtriaNet on 3D contrast-enhanced magnetic resonance imaging (CE-MRI). a**, Quantitative results for estimating the left and right atrial (LA/RA) wall thickness from the CE-MRIs measured in the mean wall thickness accuracy. **b**, Comparison of the speed of the traditional method versus our proposed AtriaNet for wall thickness estimation. The x-axis is displayed in logarithmic scale. **c**, Qualitative results for comparing the ground truth and predicted wall thickness distribution maps in 3D for three representative samples, where each row represents the same CE-MRI sample. Red regions represent thicker walls while blue regions represent thinner walls. The mean wall thickness of each chamber is also displayed along with the accuracy of wall thickness estimation. **d**, Qualitative results for the wall thickness estimations in 2D for the first sample in (**d**). The ground truths and predictionsare compared in the two rows at different slices shown in each column. Thicker regions are shown in the light red/yellow while thinner regions are shown in dark red. Selected regions with high thickness variations are enlarged in the white boxes to improve visualization. CT, crista terminalis; LAA, left atrial appendage; PM, pectinate muscles.



**Fig. 5 | Fibrosis quantification performance of AtriaNet on 3D contrast-enhanced magnetic resonance imaging (CE-MRI) and validation with low-voltage mapping of the atria. a**, Quantitative results for the left and right atrial (LA/RA) fibrosis quantification CE-MRI measured in Kolmogorov-Smirnov (KS) correlations and percentage of fibrosis error. **b**, Qualitative results in 2D comparing the ground truth and predicted fibrosis for the one representative CE-MRI at two slices. The fibrosis is shown in red. **c**,Comparison of the 3D LA ground truth and AtriaNet predicted fibrosis maps from CE-MRIs and the voltage maps showing the fibrotic distribution acquired using electro-anatomical mapping for three patients. Anatomical landmarks are outlined to reference the varying geometries acquired with the different reconstruction methods. Regions of low voltage (red) represent fibrotic tissue in the voltage maps. The correlation between the ground truth and predicted fibrosis are shown, along with the fibrotic percentage of each sample. LIPV, left inferior pulmonary vein; RIPV, right inferior pulmonary vein; RSPV, right inferior pulmonary vein.

**Methods**

### Data Acquisition and Ground Truth Segmentations

*UK Biobank:* Long-axis 2D cine MRI images were acquired from subjects in the UK Biobank prospective cohort, consisting of over 50,000 participants. Imaging was performed in Cheadle, United Kingdom, on a clinical 1.5 Tesla MAGNETOM Aera wide bore clinical scanner (Siemens Medical Solutions, Erlangen, Germany). The scanners were equipped with 48 receiver channels, an 18 channels anterior body surface coil used in combination with a 12 elements of an integrated 32 element spine coil and electrocardiogram gating for cardiac synchronization. Each participant underwent a 20-minute MRI protocol without pharmacological stressor or contrast agents. To capture the atrial function, horizontal long axis (four-chamber view) and vertical long axis (two-chamber view) balanced steady-state free precession (bSSFP) cines were acquired. The spatial resolution of the cine MRI was 1.8 × 1.8 mm2 over 50-time frames at one image slice. The four-chamber and two-chamber view images had spatial dimensions of 208 × 156 and 156 × 208 pixels, respectively. Manual image segmentation was performed by a team of eight observers overlooked by three principal investigators to standardize the segmentation procedure. The LA and RA cavities were segmented for 4860 four-chamber view images, and the LA cavity was segmented for 4860 two-chamber view images. For each MRI, two of the 50-time frames were manually annotated to capture the extremes in the atrial volume at atrial end-diastolic and end-systolic stages of the cardiac cycle. The cavity for both atrial chambers was defined by manually tracing the LA and RA blood pool. Pulmonary veins (PVs) were not included in the LA annotations. The mitral valve and tricuspid valve were delineated by a linear cut to separate the atrial and ventricular chambers.

*Utah and Waikato:* The 13,552 2D CE-MRI slices were acquired from 41 de-identified patients with atrial fibrillation at the University of Utah. In addition, 968 CE-MRI slices were acquired from 9 de-identified patients with atrial fibrillation at Waikato Hospital, Hamilton, New Zealand. The image acquisition protocols were consistent across the two centers. All scans were acquired independently. Patients underwent CE-MRI scanning to define the atrial structure prior to and/or post-catheter ablation treatment. The images were acquired with either a 1.5 Tesla Avanto or 3.0 Tesla Verio clinical whole-body scanners (Siemens Medical Solutions, Erlangen, Germany). High-resolution CE-MRIs of bi-atrial chambers were acquired approximately 20-25 minutes after the injection of 0.1 mmol/kg gadolinium contrast (Multihance, Bracco Diagnostics Inc., Princeton, NJ) using a 3D respiratory navigated, inversion recovery prepared gradient echo pulse sequence. An inversion pulse was applied every heartbeat, and fat saturation was applied immediately before data acquisition. To preserve magnetization in the image volume, the navigator was acquired immediately after the data acquisition block. Typical scan times for the CE-MRI study were between 8-15 minutes at 1.5 Tesla and 6-11 minutes using the 3 Tesla scanner (for Siemens sequences) depending on patient respiration. The spatial resolutions of all 3D CE-MRI scans were 0.625 × 0.625 × 1.25 mm³. CE-MRI scans from the University of Utah had spatial dimensions of 576 × 576 × 44 or 640 × 640 × 44 pixels, while scans from Waikato Hospital were 640 × 640 × 88 pixels. The LA and RA cavities and bi-atrial walls were manually segmented in consensus with three trained observers for each CE-MRI scan to obtain one segmentation per scan. The cavity for both atrial chambers was defined by manually tracing the LA and RA blood pool. The four PVs were included for the LA, and was limited to the PV antrum region and extended to the point where the PVs stopped narrowing. On average, the PV antra were limited to less than 10 mm extending out from the endocardial surface, or approximately three times the thickness of the LA wall. The mitral valve connecting the LA and LV and the tricuspid valve connecting the RA and RV were delineated by a 3D plane to create a smooth linear surface and separate the atrial and ventricular chambers. The cavity surfaces were then morphologically dilated and manually edited to obtain the outer boundary of the wall. The septum was manually traced to connect the walls of the LA and RA.

### Ground Truth Wall Thickness for CE-MRIs

The calculation of the wall thickness was based on a method, called a coupled partial differential equation (PDE) approach, previously proposed by Wang et al. The full set of equations can be found in the original publication45. The algorithm for computing the wall thickness produced a 3D map with a value for the thickness of the wall at every pixel inside the manually segmented LA and RA wall masks. The endocardial surface and epicardial surface were firstly manually defined. The solution to the Laplace equation *ϕ* with the endocardial and epicardial surfaces as the boundaries was obtained and then used to formulate two first-order PDEs representing a function of the distance of each pixel from the endocardial and epicardial surfaces. The PDE from the epicardial surface, *Depi*, was



where , , and

(1)

A similar PDE was then constructed for the endocardial surface. Lastly, the two distance functions were solved and summed to produce a 3D map of the AWT distribution across both atrial chambers.

### Ground Truth Fibrosis for CE-MRIs

The quantification of the atrial fibrosis was based on methods previously proposed by researchers in the University of Utah9, and was performed on both atrial chambers. The fibrotic tissues were defined as regions containing pixels with enhanced intensities within the LA and RA wall. The intensity threshold for determining the fibrotic pixels was then computed in a slice-by-slice manner by estimating the mean and standard deviation of the normal wall tissues. In this context, normal wall tissues were defined as pixels containing intensity values between the 2nd percentile to the 40th percentile of the wall. The fibrosis threshold was then calculated as 3-4 standard deviations above the mean of the normal wall tissue for each CE-MRI slice. Mathematically, this threshold calculation was written as

(2)

where *Wallnorm* was the normal wall tissue, *std* was the standard deviation, and *n* was the scaling factor to control the number of standard deviations above the mean.

### CNN Configuration for Segmentation

The pipeline consisting of two CNNs used in a sequential manner is shown in Fig.1, in which we used the CE-MRIs to demonstrate the data flow. The first CNN performed coarse segmentation on a down-sampled version of the input image to decrease computational intensity. The CE-MRIs were down-sampled four times to 144 × 144 × 44 or 144 × 144 × 88, and the cine MRI images were down-sampled two times to 104 × 78 and 78 × 104. The center of mass of the atria was calculated from the coarse segmentation in each image. A patch was then cropped centered on the center of mass, leaving out the majority of background pixels which significantly improved class imbalance for the next stage of segmentation. The size of the patch was 272 × 272 for the CE-MRIs, 80 × 96 for the four-chamber view cine MRI, and 64 × 80 for the two-chamber view cine MRI. These dimensions were selected based on the maximum size of the atrial chambers in the entire dataset, as well as ensuring it was divisible by 24 to satisfy the dimensionalities of the CNN used. The second CNN then performed regional segmentation on the ROIs cropped from the initial images. Finally, the segmentations were zero-padded to their original dimensions to obtain the final segmentation.

The same CNN was used for both stages of the two-stage pipeline and consisted of a modified U-Net architecture with additional residual connections and batch-normalization to improve the convergence. Supplemental Table. XXX shows the complete breakdown of the parameters in the network. The first half of the CNN was an encoder to learn dense features from the input through several convolutional layers of increasing depth. The convolutional layers contained 5 × 5 kernels and a stride of 1, and the number of feature maps increased from 16 to 256. At every 1 to 3 convolutional layers, residual connections were added to improve feature learning and 2 × 2 convolutions with a stride of 2 were used to progressively down sample the image by a factor of 2. The second half of the CNN was a decoder to reconstruct the image back to the original resolution for segmentation through several 5 × 5 convolutional layers of decreasing depth. The number of feature maps of the convolutions in this part of the network decreased from 128 to 32. The images were progressively up sampled by a factor of 2 with 2 × 2 deconvolutional, or transpose convolutional, layers with a stride of 2. Residual connections were also added at every 1-3 convolutional layers. In order to directly preserve high-resolution features from the input, feature forwarding connections were also used to concatenate the outputs of the convolutional layers in the encoder part to those in the decoder path at four different points along the CNN. Overall, apart from the final output layer, batch normalization (BN) and parametric rectified linear units (PReLU) were used after every convolutional layer along the entire CNN for normalization, and 25% dropout was used at every layer for regularization to decrease overfitting. The final output layer of the CNN contained a 1 × 1 convolution with a stride of 1 and a softmax activation function to predict the various classes in the data. The number of feature maps in the final convolutional layer was 4 for segmenting the CE-MRI dataset (background, LA cavity, RA cavity, bi-atrial wall), 3 for segmenting the four-chamber view cine MRI dataset (background, LA cavity, RA cavity), and 2 for segmenting the two-chamber view cine MRI dataset (background, LA cavity) to correspond with the number of classes in each data.

To further alleviate class imbalance, a multi-class dice loss function was used during training to assign higher priorities to the pixels containing the atria during prediction. The dice loss also increased the speed of convergence, by significantly reducing computational costs. The formulation of the loss function varied when training on datasets with differing numbers of classes. The general formulation for the loss function, *Ldice*, was given by

(3)

where *Nc* was the number of classes in the label. The function *Fdice(pi, gi)* represented the individual dice scores for each class from *i = 1* to *Nc*, where *pi* and *gi* represented the predicted and ground truth 2D binary masks for class *i*. *Fdice* was defined as

(4)

where *p* and *g* were of dimensions of *x* and *y*. *Nc* had a value of 3 for the CE-MRIs, 2 for the four-chamber view cine MRIs, and 1 for the two-chamber view cine MRIs.

The adaptive moment estimation (ADAM) gradient descent optimizer was used to minimize the loss function during training with a constant learning rate of 0.0001, and the exponential decay rates of the 1st and 2nd moment estimates were set to 0.9 and 0.999, respectively. To reduce the computational burden of the large images needed to be processed, all data were stored in the hierarchical data format (HDF) after pre-processing. Before each epoch, online data augmentation was used to randomly augment each data in training set with a probability of 50%. The augmentation strategies included random scaling, rotations, flipping, and elastic deformations, providing the CNN with a more diverse set of training images without increasing the memory costs associated with increasing data samples. The CNN was trained with a maximum limit of 1000 epochs, with a criterion to stop training if the accuracy on the validation set did not improve after 50 epochs. A batch size of 8 was used for the CE-MRI dataset, and 32 for the two cine MRI datasets. The training set was also shuffled for each epoch to increase randomness. After every epoch, the performance of the CNN was evaluated on the validation set with the dice score. The parameter set of the CNN which achieved the highest validation accuracy was saved and used on the testing set. The CNN was developed in TensorFlow, an open-source Python deep learning library, and TFLearn, a high-level Python API for Tensorflow. The training was performed on an Nvidia Titan V GPU with 5120 CUDA cores and 12 GB RAM. The training phase took approximately five hours. Predictions took approximately 5 seconds for each 3D CE-MRI and 0.1 second for each 2D cine MRI image.

### CNN Modifications for Wall Thickness Estimation

The CNN was modified to automatically predict the AWT for any given mask of the LA and RA walls. The ground truth AWT was computed for every manual CE-MRI segmentation with the traditional PDEs method to provide training data for the CNN. For this task, the input of the CNN was the binary LA and RA wall segmentation, and the output was the same LA and RA mask with the pixel values representing the AWT. Computing the AWT using the CNN requires the output to be continuous. The output layer was modified to contain only one feature map with a rectified linear unit (ReLU) to restrict the predictions be continuous and positive. To accommodate the presence of continuous variables, the loss function was changed to the mean squared error (MSE) between the predicted AWT values and the ground truth AWT values. During the loss calculation, the background pixels were masked to neglect the effect of the high proportion of zero-pixels which would increase the difficulty for convergence. The loss function, *LMSE*, was formulated as

(5)

where *pwall* and *gwall* represented the atrial wall pixels in the predicted and ground truth 2D AWT maps. All other parameters in the CNN remained the same as described in the previous section.

### CNN Modifications for Fibrosis Quantification

The CNN was modified to automatically predict the fibrosis directly from the CE-MRIs. The atrial fibrosis was then quantified to measure the 3D distribution and percentage of diseased versus healthy tissue in each patient. This was performed on the bi-atrial wall of the predicted segmentations which defined as the LA and RA walls. To allow for the additional labels of LA fibrosis and RA fibrosis, the output of the CNN was modified from its segmentation version to accommodate four output labels for the background, non-fibrotic bi-atrial wall, LA fibrosis, and RA fibrosis. The same loss function as in Eq. 3 was used with the new labels. All other parameters in the CNN remained the same as described in the previous section.

### Technical Evaluation Metrics

Technical measures for segmentation included the mean squared error (MSE), dice score, STSD, sensitivity and specificity. The MSE was used to measure the first CNN in the two-stage pipeline, by calculating the accuracy of the location of the center of mass of the atria detected for ROI detection. It was given by

(6)

for the ground truth co-ordinates *(x, y)* and the predicted coordinates *(x’, y’)* of the center of mass in each slice of the 2D Cine MRI and 3D CE-MRIs. The dice score is the most commonly used metric for evaluating image segmentation accuracy, and was given by

(7)

for a prediction mask, *A*, and a ground truth, *B*. The STSD measures the average distance error between the surfaces of the predicted atrial mask, *A*, and the ground truth, *B*, and was given by

(8)

where *nA* is the number of pixels in *A*, *nB* is the number of pixels in *B*, and *p* and *p’* describe all points in *A* and *B*.

The technical measure used for wall thickness estimation was the average pixel-wise error which quantified the error of each predicted wall thickness value with

(9)

to show the absolute error of the CNN.

The technical measure used for fibrosis quantification was the Kolmogorov-Smirnov (KS) score25 which measures the correlation between the ground truth and predicted spatial distributions of the fibrosis in the LA and RA walls. The KS score was defined as

(10)

where *FA­(x)* and *FB(x)* were the cumulative frequency distributions of the prediction, *A*, and ground truth, *B*, LA fibrosis co-ordinates in 3D. The KS score produced values between 0 to 1, where 1 represented the prediction and ground truth containing identical distributions.

### Biological Evaluation Metrics

Biological measures were also used to validate the clinical usability of the CNN for segmentation. It included the atrial anterior-posterior diameter error and atrial volume error between the predictions and ground truths which are the two most commonly used clinical biomarkers. These measures were converted into percentage accuracies for ease of understanding. The atrial diameterwas calculated by measuring the maximum distance from the anterior to the posterior sides of the atrial chambers in 2D for the cine-MRIs and in 3D for the CE-MRIs. The diameter error was then measured as the percentage difference between the LA and RA diameters measured from the predicted and ground truth segmentation masks. The atrial volume was calculated by summing all pixels in the segmentation mask defining the LA and RA chamber cavities. Similar to the diameter, the volume error was measured as a percentage difference between the predicted and ground truth LA and RA volumes. When evaluating the cine-MRIs, the atrial areas were used due to the images being 2D.

The biological measure used for wall thickness estimation was the percentage difference between the mean LA and RA wall thicknesses of the predicted and ground truth maps. Lastly, the fibrosis error was measured as the difference between the fibrotic percentage as a proportion of the LA and RA wall pixels in the predictions and ground truths.

**Data Availability**

The CE-MRI data is openly accessible on https://atriaseg2018.cardiacatlas.org/.

**Code Availability**

Full implementation of the convolutional neural network used will be provided once the paper is published.

**References**

1. Nabel, E.G. Cardiovascular disease. *New England Journal of Medicine* **349**, 60-72 (2003).

2. Medrano-Gracia, P., Cowan, B.R., Suinesiaputra, A. & Young, A.A. Challenges of cardiac image analysis in large-scale population-based studies. *Current cardiology reports* **17**, 9 (2015).

3. Oakes, R.S.*, et al.* Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. *Circulation* **119**, 1758-1767 (2009).

4. Csepe, T.A.*, et al.* Novel application of 3D contrast-enhanced CMR to define fibrotic structure of the human sinoatrial node in vivo. *European Heart Journal-Cardiovascular Imaging*, jew304 (2017).

5. Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B. & Neubauer, S. The role of cardiovascular magnetic resonance imaging in heart failure. *Journal of the American College of Cardiology* **54**, 1407-1424 (2009).

6. Tsang, T.S.*, et al.* Left atrial volume: important risk marker of incident atrial fibrillation in 1655 older men and women. in *Mayo Clinic Proceedings*, Vol. 76 467-475 (Elsevier, 2001).

7. Bishop, M.*, et al.* Three-dimensional atrial wall thickness maps to inform catheter ablation procedures for atrial fibrillation. *Europace* **18**, 376-383 (2016).

8. Matsuda, Y.*, et al.* Importance of left atrial function in patients with myocardial infarction. *Circulation* **67**, 566-571 (1983).

9. McGann, C.*, et al.* Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI. *Circulation: Arrhythmia and Electrophysiology* **7**, 23-30 (2014).

10. Maeder, M.T. & Kaye, D.M. Heart failure with normal left ventricular ejection fraction. *Journal of the American College of Cardiology* **53**, 905-918 (2009).

11. Higuchi, K.*, et al.* The Spatial Distribution of Late Gadolinium Enhancement of Left Atrial MRI in Patients With Atrial Fibrillation. *JACC: Clinical Electrophysiology* (2017).

12. Bai, W.*, et al.* A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature Medicine* **26**, 1654-1662 (2020).

13. Bernard, O.*, et al.* Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE transactions on medical imaging* **37**, 2514-2525 (2018).

14. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical image computing and computer-assisted intervention* 234-241 (Springer, 2015).

15. Zhuang, X.*, et al.* Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical image analysis* **58**, 101537 (2019).

16. Xiong, Z.*, et al.* A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical Image Analysis* **67**, 101832 (2021).

17. Yang, G., Hua, T. & Xue, W. Left ventricle full quantification challenge. in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent* (2019).

18. Xiong, Z.*, et al.* Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network. *IEEE transactions on medical imaging* **38**, 515-524 (2018).

19. Petersen, S.E.*, et al.* Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *Journal of Cardiovascular Magnetic Resonance* **19**, 1-19 (2017).

20. Xiong, Z.*, et al.* Fully automatic 3D bi-atria segmentation from late gadolinium-enhanced MRIs using double convolutional neural networks. in *International Workshop on Statistical Atlases and Computational Models of the Heart* 63-71 (Springer, 2019).

21. Foo, F.S.*, et al.* Regional variation in cardiac implantable electronic device implants trends in New Zealand over the past decade (ANZACS‐QI 54). *Internal Medicine Journal* (2020).

22. Bai, W.*, et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance* **20**, 65 (2018).

23. Zhao, J.*, et al.* Three‐dimensional integrated functional, structural, and computational mapping to define the structural “fingerprints” of heart‐specific atrial fibrillation drivers in human heart ex vivo. *Journal of the American Heart Association* **6**, e005922 (2017).

24. Song, J.-S.*, et al.* Role of atrial wall thickness in wave-dynamics of atrial fibrillation. *PloS one* **12**, e0182174 (2017).

25. Weiss, M.S. Modification of the Kolmogorov-Smirnov statistic for use with correlated data. *Journal of the American Statistical Association* **73**, 872-875 (1978).

26. Benito, E.M.*, et al.* Left atrial fibrosis quantification by late gadolinium-enhanced magnetic resonance: a new method to standardize the thresholds for reproducibility. *Ep Europace* **19**, 1272-1279 (2017).

27. Pontecorboli, G.*, et al.* Use of delayed-enhancement magnetic resonance imaging for fibrosis detection in the atria: a review. *EP Europace* **19**, 180-189 (2017).

28. Sramko, M.*, et al.* Clinical value of assessment of left atrial late gadolinium enhancement in patients undergoing ablation of atrial fibrillation. *International journal of cardiology* **179**, 351-357 (2015).

29. Harrison, J.L.*, et al.* Repeat left atrial catheter ablation: cardiac magnetic resonance prediction of endocardial voltage and gaps in ablation lesion sets. *Circulation: Arrhythmia and Electrophysiology* **8**, 270-278 (2015).

30. Zhuang, X.*, et al.* Cardiac segmentation on late gadolinium enhancement MRI: a benchmark study from multi-sequence cardiac MR segmentation challenge. *arXiv preprint arXiv:2006.12434* (2020).

31. Li, C.*, et al.* Attention based hierarchical aggregation network for 3D left atrial segmentation. in *International Workshop on Statistical Atlases and Computational Models of the Heart* 255-264 (Springer, 2018).

32. Isensee, F.*, et al.* Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. in *International workshop on statistical atlases and computational models of the heart* 120-129 (Springer, 2017).

33. Dittrich, H.C.*, et al.* Left atrial diameter in nonvalvular atrial fibrillation: an echocardiographic study. *American heart journal* **137**, 494-499 (1999).

34. Tobon-Gomez, C.*, et al.* Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE transactions on medical imaging* **34**, 1460-1473 (2015).

35. Burstein, B. & Nattel, S. Atrial fibrosis: mechanisms and clinical relevance in atrial fibrillation. *Journal of the American College of Cardiology* **51**, 802-809 (2008).

36. Yamazaki, M.*, et al.* Heterogeneous atrial wall thickness and stretch promote scroll waves anchoring during atrial fibrillation. *Cardiovascular research* **94**, 48-57 (2012).

37. Yang, G.*, et al.* Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. *Future Generation Computer Systems* **107**, 215-228 (2020).

38. Kwon, O.-S.*, et al.* Accuracy and clinical feasibility of 3D-myocardial thickness map measured by cardiac computed tomogram. *International Journal of Arrhythmia* **21**, 1-11 (2020).

39. Kamali, R.*, et al.* Reproducibility of clinical late gadolinium enhancement magnetic resonance imaging in detecting left atrial scar after atrial fibrillation ablation. *Journal of Cardiovascular Electrophysiology* **31**, 2824-2832 (2020).

40. Huang, Q.*, et al.* Measure Anatomical Thickness from Cardiac MRI with Deep Neural Networks. in *International Workshop on Statistical Atlases and Computational Models of the Heart* 44-55 (Springer, 2020).

41. Karim, R.*, et al.* Algorithms for left atrial wall segmentation and thickness–Evaluation on an open-source CT and MRI image database. *Medical image analysis* **50**, 36-53 (2018).

42. Goodfellow, I.J., Vinyals, O. & Saxe, A.M. Qualitatively characterizing neural network optimization problems. *arXiv preprint arXiv:1412.6544* (2014).

43. Johnson, J.M. & Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *Journal of Big Data* **6**, 1-54 (2019).

44. Kettering, K.*, et al.* Catheter ablation of atrial fibrillation using the Navx-/Ensite-system and a CT-/MRI-guided approach. *Clinical research in cardiology* **98**, 285-296 (2009).

45. Wang, Y.*, et al.* A robust computational framework for estimating 3D Bi-Atrial chamber wall thickness. *Computers in biology and medicine* **114**, 103444 (2019).

**Acknowledgements**

This work was funded by the Heath Research Council of New Zealand and Heart Foundation New Zealand. We are grateful for Nvidia for donating a Titan-V GPU which aided our algorithm development. We would like to acknowledge the NIH/NIGMS Center for Integrative Biomedical Computing (CIBC) at the University of Utah, United States. This research has also been conducted using data from UK Biobank, a major biomedical databse: [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk/" \t "_blank)

**Ethics declarations**

This study complies with the Declaration of Helsinki; the work was covered by the ethical approval for UK Biobank studies from the NHS National Research Ethics Service on 17th June 2011 (Ref 11/NW/0382) and extended on 18 June 2021 (Ref 21/NW/0157) with written informed consent obtained from all participants. No other ethics declaration applies.