

# Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced MRI

Guang Yang<sup>a)</sup>

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK  
National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK*

Xiahai Zhuang<sup>a)</sup>

*School of Data Science, Fudan University, Shanghai 201203, China*

Habib Khan, Shouvik Haldar, and Eva Nyktari

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK*

Lei Li

*Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Ricardo Wage

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK*

Xujiong Ye

*School of Computer Science, University of Lincoln, Lincoln LN6 7TS, UK*

Greg Slabaugh

*Department of Computer Science, City University London, London EC1V 0HB, UK*

Raad Mohiaddin

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK  
National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK*

Tom Wong

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK*

Jennifer Keegan\* and David Firmin\*

*Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK  
National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK*

(Received 12 October 2017; revised 1 February 2018; accepted for publication 17 February 2018; published 15 March 2018)

**Purpose:** Atrial fibrillation (AF) is the most common heart rhythm disorder and causes considerable morbidity and mortality, resulting in a large public health burden that is increasing as the population ages. It is associated with atrial fibrosis, the amount and distribution of which can be used to stratify patients and to guide subsequent electrophysiology ablation treatment. Atrial fibrosis may be assessed noninvasively using late gadolinium-enhanced (LGE) magnetic resonance imaging (MRI) where scar tissue is visualized as a region of signal enhancement. However, manual segmentation of the heart chambers and of the atrial scar tissue is time consuming and subject to interoperator variability, particularly as image quality in AF is often poor. In this study, we propose a novel fully automatic pipeline to achieve accurate and objective segmentation of the heart (from MRI Roadmap data) and of scar tissue within the heart (from LGE MRI data) acquired in patients with AF.

**Methods:** Our fully automatic pipeline uniquely combines: (a) a multiatlas-based whole heart segmentation (MA-WHS) to determine the cardiac anatomy from an MRI Roadmap acquisition which is then mapped to LGE MRI, and (b) a super-pixel and supervised learning based approach to delineate the distribution and extent of atrial scarring in LGE MRI. We compared the accuracy of the automatic analysis to manual ground truth segmentations in 37 patients with persistent long-standing AF.

**Results:** Both our MA-WHS and atrial scarring segmentations showed accurate delineations of cardiac anatomy (mean Dice = 89%) and atrial scarring (mean Dice = 79%), respectively, compared to the established ground truth from manual segmentation. In addition, compared to the ground truth, we obtained 88% segmentation accuracy, with 90% sensitivity and 79% specificity. Receiver operating characteristic analysis achieved an average area under the curve of 0.91.

**Conclusion:** Compared with previously studied methods with manual interventions, our innovative pipeline demonstrated comparable results, but was computed fully automatically. The proposed segmentation methods allow LGE MRI to be used as an objective assessment tool for localization, visualization, and quantitation of atrial scarring and to guide ablation treatment. © 2018 The Authors.

*Medical Physics* published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine. [https://doi.org/10.1002/mp.12832]

[The copyright line for this article was changed on May 18, 2018, after original online publication].

Key words: atrial fibrillation cardiovascular magnetic resonance imaging, late gadolinium-enhanced MRI, medical image segmentation, whole heart segmentation

### Abbreviation

AF	atrial fibrillation
LA	left atrium
PV	pulmonary veins
SVM	support vector machine
CV	cross-validation
ROI	region of interest
LGE-MRI	late gadolinium-enhanced magnetic resonance imaging

## 1. INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia of clinical significance, resulting in a large public health burden that is increasing as the population ages. It occurs when chaotic and disorganized electrical activity develops in the atria, causing muscle cells to contract irregularly and rapidly. It is associated with structural remodeling, including fibrotic changes in the left atrium<sup>1</sup> and can cause increased morbidity, especially stroke and heart failure. It also results in poor mental health, dementia, and increased mortality, costing the NHS £2.2 billion per year.<sup>2</sup>

The electrical impulses that trigger AF frequently originate in the pulmonary veins (PV). Radio frequency ablation treatment aims to eliminate AF by electrically isolating the PV. However, the success rate for a single catheter ablation procedure is just 30–50% at 5-yr follow-up<sup>3,4</sup> and multiple ablations are frequently required. Accurate knowledge of native and previous ablation scarring could be used to stratify patients and guide ablation treatment, ultimately reducing the need for repeat procedures.<sup>5</sup>

The current clinical gold standard for assessment of atrial scarring is electro-anatomical mapping (EAM), performed during an electrophysiological (EP) study.<sup>6</sup> However, this is an invasive technique which uses ionizing radiation and the accuracy is suboptimal, with reported errors of up to 10 mm in the localization of scar tissue.<sup>7,8</sup>

Late gadolinium enhancement (LGE) magnetic resonance imaging (MRI) is an established noninvasive technique for detecting myocardial scar tissue.<sup>9</sup> With this technique, healthy and scar tissues are differentiated by their altered wash-in and wash-out contrast agent kinetics, which result in scar tissue being seen as a region of enhanced or high signal intensity while healthy tissue is “nulled.” Atrial 3D LGE MRI has been used to assess patient suitability for AF ablation by identifying potential nonresponders,<sup>10–16</sup> and to define the most appropriate ablation approach.<sup>12,13,17</sup> In addition, visualization and quantitation of native and postablation atrial scarring derived from LGE MRI has been used to guide

initial and follow-up ablation procedures.<sup>13,14,18–21</sup> Histopathological studies in pigs have validated LGE MRI for the characterization of AF ablation-induced wall injury.<sup>22</sup>

Visualization and quantification of atrial scarring requires objective, robust, and accurate segmentation of the enhanced scar regions from the LGE MRI images. Essentially, there are two segmentations required: one showing the cardiac anatomy (geometry), particularly the LA wall and PV, the other delineating the enhanced scar regions. The former segmentation is required to rule out confounding enhanced tissues from other parts of the heart, for example, the mitral valve and aorta, or the enhancement from nonheart structures while the latter is a prerequisite for visualization and quantitation. Segmentation of the atrial scarring from LGE MRI images is a very challenging problem. Firstly, the LA wall is very thin and scarring is hard to distinguish even by experienced expert cardiologists specialized in cardiac MRI. Secondly, residual respiratory motion, heart rate variability, low signal-to-noise ratio (SNR), and contrast agent wash-out during the long acquisition (current scanning time  $\approx 10$  min) frequently result in image quality being poor. Artifactual enhanced signal from surrounding tissues may also result in a large number of false positives.

A grand challenge for evaluation and benchmarking of various atrial scarring segmentation methods has shown promising results<sup>5</sup> although most have relied on manual segmentation of the LA wall and PV. This has several drawbacks: (a) it is a time-consuming task; (b) there are intra- and interobserver variations; and (c) it is less reproducible for a multicenter and multicenter study. Moreover, a number of studies have assumed a fixed thickness of the LA wall although there is no evidence that this is the case. Depending on the actual wall thickness, subsequent reorientation and interpolation of the MR images results in varying partial volume effects, which affect the apparent thickness of the LA wall. Inaccurate manual segmentation of the LA wall and PV can further complicate the delineation of the atrial scarring and its quantitation can be error prone.

The LA and PV would ideally be segmented from the cardiac and respiratory-gated LGE MRI dataset. However, this is difficult as the inversion magnetization preparation used reduces the blood pool signal and normal myocardium is nulled. Other options are to segment them from a separately acquired breath-hold magnetic resonance angiogram (MRA) study<sup>15,23,24</sup> or from a respiratory and cardiac gated 3D balanced steady-state-free precession (b-SSFP) “Roadmap” study.<sup>25</sup> While breath-hold MRA shows the LA and PV with high contrast, these acquisitions are generally ungated and acquired in an inspiratory breath-hold. The anatomy extracted can therefore be highly deformed compared to that in the LGE MRI study. Although the gated 3D Roadmap acquisition takes longer to acquire, it is in the same respiratory phase

as the LGE MRI and the extracted anatomy can be better matched. Cardiac anatomy has previously been defined by atlas-based segmentation of MRA<sup>24</sup> and by using a statistical shape model<sup>25</sup> on 3D Roadmap data. Table I provides a summary of previously published methods on atrial scarring segmentation using LGE MRI and our proposed method.

We propose a novel fully automatic segmentation and objective assessment of atrial scarring for long-standing persistent AF patients scanned by LGE MRI. The LA chamber and PV are defined using a multiatlas-based whole heart segmentation (MA-WHS) method on b-SSFP Roadmap MRI images. LA and PV geometry is resolved by mapping the segmented Roadmap anatomy to LGE MRI using the DICOM header data and is further refined by affine and nonrigid registration steps. LGE MRI images are segmented by a novel Simple Linear Iterative Clustering (SLIC)-based super-pixels method. A support vector machine (SVM)-based supervised classification is then applied to segment the atrial scarring within the segmented LA and PV geometry. In this study, two validation steps have been performed — one for the LA chamber and PV segmentation and one for the atrial scarring segmentation — both against established ground truth from manual segmentations by experienced expert cardiologists specialized in cardiac MRI.

## 2. MATERIALS AND METHODS

### 2.A. Data acquisition

Cardiac MR data were acquired on a Siemens Magnetom Avanto 1.5T scanner (Siemens Medical Systems, Erlangen, Germany).

Transverse navigator-gated 3D LGE MRI<sup>10,26,27</sup> was performed using an inversion prepared segmented gradient echo

sequence (TE/TR 2.2 ms/5.2 ms) 15 min after gadolinium (Gd) administration (Gadovist — gadobutrol, 0.1 mmol/kg body weight, Bayer-Schering, Berlin, Germany).<sup>28</sup> The inversion time was set to null the signal from normal myocardium. Detailed scanning parameters are: 30–34 slices at  $1.5 \times 1.5 \times 4 \text{ mm}^3$ , reconstructed to 60–68 slices at  $0.75 \times 0.75 \times 2 \text{ mm}^3$ , field-of-view  $380 \times 380 \text{ mm}^2$ , acceleration factor of 2 using generalized autocalibrating partially parallel acquisition (GRAPPA), acquisition window 125 ms positioned within the subject-specific rest period, single R-wave gating, chemical shift fat suppression, flip angle  $20^\circ$ . Data were acquired during free-breathing using a crossed-pairs navigator positioned over the dome of the right hemi-diaphragm with navigator acceptance window size of 5 mm and CLAWS respiratory motion control.<sup>29</sup> The nominal acquisition duration was 204–232 cardiac cycles assuming 100% respiratory efficiency.

Prior to contrast agent administration, coronal navigator-gated 3D b-SSFP (TE/TR 1/2.3 ms) Roadmap data were acquired with the following parameters: 80 slices at  $1.6 \times 1.6 \times 3.2 \text{ mm}^3$ , reconstructed to 160 slices at  $0.8 \times 0.8 \times 1.6 \text{ mm}^3$ , field-of-view  $380 \times 380 \text{ mm}^2$ , acceleration factor of 2 using GRAPPA, partial Fourier 6/8, acquisition window 125 ms positioned within the subject-specific rest period, chemical shift fat suppression, flip angle  $70^\circ$ . Off resonant blood from the lungs arriving in the LA and PV can result in signal loss,<sup>30</sup> which in our application, is minimized by using the shortest TE/TR possible. This was achieved by using nonselective RF excitation.<sup>31</sup> Data were acquired during free-breathing using a crossed-pairs navigator positioned over the dome of the right hemi-diaphragm with navigator acceptance window size of 5 mm and CLAWS respiratory motion control.<sup>29</sup> The nominal acquisition duration was 241 cardiac cycles assuming 100% respiratory efficiency.

TABLE I. Summary of the previously published methods for atrial scarring segmentation and our proposed method.

References	Subjects (number)	Cardiac anatomy segmentation (modality)	Atrial scarring segmentation	Evaluation of atrial scarring segmentation (results: mean $\pm$ std)
Oakes et al. <sup>10</sup>	Human (81)	Manual segmentation of LA wall (LGE MRI)	2–4 SD	Atrial scarring percentage ( $8 \pm 4$ , $21 \pm 6$ , $50 \pm 15$ ) <sup>a</sup>
Knowles et al. <sup>23</sup>	Human (7)	Semi-automatic thresholding and region growing (MRA)	Maximum intensity projection	Atrial scarring percentage ( $31 \pm 10$ ) <sup>b</sup>
Perry et al. (2012) <sup>68</sup>	Human (34)	Manual segmentation of LA wall (LGE MRI)	k-means Clustering	Dice ( $81 \pm 11$ , ground truth by manually selected thresholds)
Ravanelli et al. <sup>15</sup>	Human (10)	Manual segmentation of LA and PV in 3D (MRA)	4 SD	Dice ( $60 \pm 21$ ground truth by a semi-automatic approach) <sup>c</sup>
Karim et al. <sup>25</sup>	Human (15)	Statistical shape model with manual correction (b-SSFP)	Graph cuts	Dice, ROC, and total scar volume <sup>d</sup>
Tao et al. <sup>24</sup>	Human (46)	Automatic atlas-based method with level set refinement (MRA)	Maximum intensity projection	Qualitative visualization (N/A)
Proposed method	Human (37)	Fully automated multiatlas whole heart segmentation (b-SSFP)	Super-pixel and SVM	Multiple quantitative metrics (Dice: $79 \pm 5$ )

<sup>a</sup>Results (%) for mild ( $n = 43$ ), moderate ( $n = 30$ ), and extensive ( $n = 8$ ) enhancement cases.

<sup>b</sup>Moderate and extensive enhancement cases.

<sup>c</sup>The Dice score was calculated for an automated atrial scarring segmentation. The method was also evaluated using Bland–Altman analysis of the atrial scarring percentage (after skeletonization) obtained from LGE MRI and EAM.

<sup>d</sup>Multiple Dice scores were calculated for various experimental settings, and they were reported by plotting the median Dice scores (around 80) with the minimum and the maximum.

### 2.B. Patients

In agreement with the local regional ethics committee, cardiac MRI was performed in long-standing persistent AF patients between 2011 and 2013. Gadolinium administration is contraindicated in patients with severe kidney disease due to an increased, but still rare, risk of developing nephrogenic systemic fibrosis. Consequently, LGE MRI was not performed in any patient with 30 ml/min/1.73 m<sup>2</sup>.

The image quality of each dataset was scored by a senior cardiac MRI physician on a Likert-type scale — 0 (non-diagnostic), 1 (poor), 2 (fair), 3 (good), and 4 (very good) — depending on the level of SNR, appropriate TI, and the existence of navigator beam and ghost artifacts. Thirty-seven cases with image quality greater or equal to 2 were retrospectively entered into this study including 11 preablation (included ~65% of preablation cases) and 26 postablation scans (included ~92% of post-ablation cases).

### 2.C. Multiatlas whole heart segmentation (MA-WHS)

A multiatlas approach<sup>32,33</sup> was developed to derive the whole heart segmentation of the Roadmap acquisition, which was then mapped to LGE MRI [Fig. 1(a)]. This segmentation consists of two major steps: (a) atlas propagation based on image registration algorithms and (b) label fusion from multi-atlas-propagated segmentation results.

First we obtained 30 MRI Roadmap studies from the Left Atrium Segmentation Grand Challenge organized by King’s College London<sup>34</sup> together with manual segmentations of the left atrium, pulmonary veins, and appendages. In these, we further labeled the right and left ventricles, the right atrium, the aorta and the pulmonary artery, to generate 30 whole

heart atlases. These 30 MRI Roadmap studies were employed only for building an independent multiatlas dataset, which is then used for segmenting our Roadmap studies that are linked with the LGE MRI scans for the AF patients.

Let  $I$  be the target image to be segmented,  $\{(A_a, L_a) | a = 1, \dots, N\}$  be the set of atlases, where  $N = 30$ ,  $A_a$  and  $L_a$  are, respectively, the intensity image and corresponding segmentation label image of the  $a$ th atlas. For each atlas, MA-WHS performs an atlas-to-target registration, by maximizing the similarity between the images, to derive the set of warped atlases,

$$T_a = \arg \min_{T_a} \text{ImageSimilarity}(I, A_a), \text{ and } \begin{cases} A_a = T_a(A_a) \\ L_a = T_a(L_a) \end{cases} \quad (1)$$

in which  $T_a$  is the resulting transformation of the registration and  $\{(A_a, L_a) | a = 1, \dots, N\}$  are, respectively, the warped atlas intensity image and corresponding segmentation result. Here, we employ the hierarchical registration for segmentation propagation, which was specifically designed for the whole heart MRI images and consists of three steps, namely the global affine registration for localization of the whole heart, the local affine registration for the initialization of the substructures, and the fully deformable registration for local detail refinement.<sup>35</sup> Image similarity metrics evaluate how similar the atlas and target image are. In this work, we propose to use the spatially encoded mutual information (SEMI) method, which has been shown to be robust against intensity nonuniformity and different intensity contrast,<sup>36</sup> that is

$$\text{ImageSimilarity}(I, A_a) = \{S_1, \dots, S_{n_s}\} \quad (2)$$

where  $\{S_1, \dots, S_{n_s}\}$  are the SEMI and computed based on the spatially encoded joint histogram,

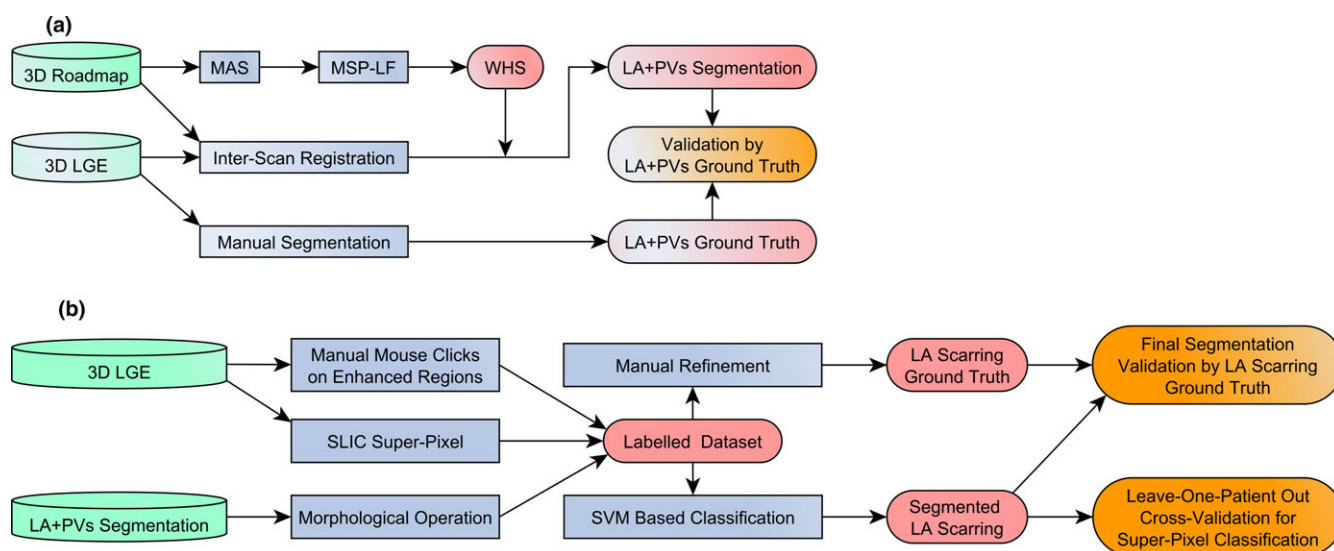


FIG. 1. (a) Flowchart of the LA + PV segmentation via MA-WHS and its validation. (b) Flowchart of the fully automatic atrial scarring segmentation including atrial scarring ground truth construction, super-pixel and SVM classification-based segmentation and leave-one-patient-out cross-validation. Abbreviations: LA + PV, left atrium and pulmonary veins; MAS, multiatlas propagation-based segmentation; MSP-LF, multiscale patch-based label fusion; WHS, whole heart segmentation. [Color figure can be viewed at wileyonlinelibrary.com]



$$H_s(I, A_a) = \sum_{x \in \Omega} w_1(I(x)) w_2(A_a(x)) W_s(x). \quad (3)$$

Here,  $w_1(I(x))$  and  $w_2(A_a(x))$  are Parzen window estimation and  $W_s(x)$  is a weighting function to encode the spatial information.<sup>36</sup>

After the multiatlas propagation, a label fusion algorithm is required to generate one final segmentation of the LA from the 30 propagated results,

$$L_I = \text{LabelFusion}(\{(A_1, L_1) \dots (A_N, L_N)\}). \quad (4)$$

The label fusion decides how to combine the multiple classification results into one labeling result. Since the atlases can produce segmentations with dramatically different accuracy at different locations, it should evaluate the performance of each atlas locally and assigns different weights for the atlases at each pixel of the target image in decision fusion.

The recent literatures have many new methods<sup>37–44</sup> on improving multiatlas segmentation using sophisticatedly designed algorithms, which generally need to evaluate local similarity between patches from the atlases and the target image for local weighted label fusion,

$$L_I(x) = \operatorname{argmax}_{l \in \{l_{bk}, l_{la}\}} \sum_a w_a(S(I, A_a, x)) \delta(L_a(x), l), \quad (5)$$

in which  $l_{bk}$  and  $l_{la}$  indicate the labels of the background and left atrium, respectively, and the local weight  $w_a(\cdot) \propto S(\cdot)$  is determined by the local similarity  $S(\cdot)$  between the target image and the atlas.  $\delta(a, b)$  is the Kronecker delta function which returns 1 when  $a = b$  and returns 0 otherwise.

For the LA segmentation, we propose to use the multiscale patch-based label fusion (MSP-LF). This is because the intensity distribution of the blood pool in the LA is almost identical to that of the blood pool in the other chambers and great vessels. The multiscale space theory can handle different level information within a small patch and has been applied to feature extraction/detection and image matching.<sup>33,44–50</sup> The patches we compute from different scale spaces can represent the different levels of structural information, with low scale capturing local fine structure and high scale suppressing fine structure but providing global structural information of the image. This is different from the conventional patch-based methods, which only compute the local structural information within the patch. To avoid increasing the computational complexity, we adopt the multiresolution implementation and couple it with the MSP where the high-scale patch can be efficiently computed using a low-resolution image space. The local similarity between two images using the MSP measure is computed, as follows,

$$S_{\text{msp}}(I, A_a, x) = \sum_s S(I^{(s)}, A_a^{(s)}, x) \quad (6)$$

where  $I^{(s)} = I * \text{Gaussian}(0, \sigma_s)$  is the target image from  $s$  scale space which is computed from the convolution of the target image with Gaussian kernel function with scale  $s$ . Here, we compute the local similarity in multiscale image using the conditional probability of the images,

$$S(I^{(s)}, A_a^{(s)}, x) = p(i_x | j_x) = \frac{p(i_x, j_x)}{p(j_x)}. \quad (7)$$

where  $i_x = I^{(s)}(x)$  and  $j_x = A_a^{(s)}(x)$  and the conditional image probability is obtained from the joint and marginal image probability which can be calculated using the Parzen window estimation.<sup>51</sup>

For each patient, the Roadmap dataset was then registered to the LGE MRI dataset using the DICOM header data, and then refined by affine and nonrigid registration steps.<sup>36</sup> The resulting transformation was applied to the MA-WHS-derived cardiac anatomy to define the endocardial LA boundary and PV on the LGE MRI dataset for each patient. It is of note that both our Roadmap and LGE MRI data were cardiac gated and acquired in end expiration, which minimizes significant shape deformation between the two.

## 2.D. Atrial scarring segmentation

### 2.D.1. Oversegmentation by simple linear iterative clustering (SLIC)-based super-pixels

We used a simple linear iterative clustering (SLIC)-based super-pixel method<sup>52</sup> to oversegment LGE MRI images in order to separate potential enhanced atrial scarring regions from other tissues [Fig. 1(b)]. Super-pixel algorithms group pixels into perceptually meaningful patches with similar size, which can be used to replace the regular pixel grid. Consequently, the derived super-pixel patches can capture and mitigate image redundancy, and therefore provide a significant primitive from which image features can be calculated effectively and efficiently. In summary, super-pixel methods have been proven to have following benefits: (a) super-pixels can adhere well to perceptually meaningful object boundaries in images; (b) super-pixels can reduce computational complexity of extracting image features; (c) for segmentation applications, super-pixels can improve the performance while reducing the computation time.<sup>53</sup> In this study, we proposed to use a SLIC-based super-pixel method, which has been successfully applied to solve various medical image analysis problems.<sup>54,55</sup> It has also demonstrated better segmentation accuracy and superior adherence to object boundaries, and it is faster and more memory efficient compared to other state-of-the-art super-pixels methods.<sup>52</sup> Based on local k-means clustering, the SLIC method iteratively groups pixels into super-pixels. The clustering proximity is estimated in both intensity and spatial domains that is

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2}, \quad (8)$$

in which  $d_c = \sqrt{(I_j - I_i)^2}$  measures the pixel intensity difference of a gray scale image and  $d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$  describes the spatial distance between each pixel and the geometric center of the super-pixel. SLIC is initialized by sampling the target slice of the LGE MRI image into a regular grid

space with grid interval of  $S$  pixels. To speed up the iteration, SLIC limits the size of search region of similar pixels to  $2S \times 2S$  around the super-pixel center (namely local k-means clustering). In addition, parameter  $m$  balances the weighting between intensity similarity  $d_c$  and spatial proximity  $d_s$ . In this study, we initialized  $S$  to 4 pixels that is  $2.8 \times 2.8 \text{ mm}^2$  considering the LA wall thickness is approximately 3 mm,<sup>56,57</sup> and also take into account that the super-pixel size is still large enough to extract statistics of the grouped pixel intensities. In addition,  $m$  was chosen by visual inspection of the oversegmented results, and it was fixed when the super-pixel results adhered well with the LA wall boundary.

### 2.D.2. Support vector machines (svm) based classification

After SLIC segmentation, we use Support Vector Machines (SVM) to classify the oversegmented super-pixels into enhanced atrial scarring regions and nonenhanced tissues. SVM provide a powerful technique for supervised binary classification<sup>58</sup> (refer to Supporting Material Appendix S1).

In order to train the SVM classifier, we built a training dataset containing enhanced and nonenhanced super-pixel patches. This has been done by (a) an experienced expert cardiologist specialized in cardiac MRI performing manual mouse clicks to select the enhanced scar regions; (b) combining the mouse clicks and SLIC segmentation to label the enhanced super-pixels; (c) applying morphological dilation (3 mm) to the segmented endocardial LA boundary and PV from MA-WHS to extract the LA wall and PV; (d) finding the overlapped regions of the LA wall and PV and the labeled enhanced super-pixels, and (e) labeling the other super-pixels overlapped with LA wall and PV as nonenhancement. Details of each step are given as following:

- (1) Manual mouse clicks: Instead of manually drawing the boundaries of the enhanced atrial scarring regions, we asked an experienced cardiologist specialized in cardiac MRI to perform manual mouse clicks on the LGE MRI images to label the regions that they believed to be enhanced (i.e., atrial scarring tissue). This is because manual boundary drawing of enhancement on the thin LA wall is a very challenging task and subject to large inter- and intraobserver variances. Mouse clicks on the enhancement regions are much easier and much more efficient. The manual mouse clicks were done on the original LGE MRI images without the super-pixel grid overlaid. This is because: (a) the mouse clicks will not be biased by super-pixel patches and (b) the super-pixel grid may reduce the visibility of the enhancement on LGE MRI images.
- (2) The coordinates of the mouse clicks were used to select the enhanced super-pixels. Because the cardiologist performed the mouse clicks on the original LGE MRI images without having prior knowledge about the super-pixels, we asked the cardiologist to have relatively dense mouse clicks. These mouse clicks will

ensure all the enhanced regions can be included, but only one mouse click will be taken into account if multiple clicks dwell in the same super-pixel.

- (3) The endocardial LA boundary and PV were extracted using our MA-WHS method. We then applied a morphological dilation to extract the LA wall and PV assuming that the thickness of LA wall is 3 mm. The blood pool regions were extracted by a morphological erosion (5 mm) from the endocardial LA boundary, and the pixel intensities were normalized according to the mean and standard deviation of the blood pool intensities.<sup>5</sup>
- (4) We masked the selected enhanced super-pixels [derived from step (2)] using the LA wall and PV segmentation. Only the super-pixels having a defined overlap with the LA wall and PV segmentation were selected as enhancement for building the training data (overlapping ratio was set to  $\geq 20\%$ ). Other super-pixels (overlapping ratio  $< 20\%$ ) were discarded as they were considered as enhancement from other substructures of the heart (such as the mitral valve and aorta) but not enhancement of the LA wall and PV. *It is of note that although we assumed that the LA wall thickness is 3 mm, our enhanced super-pixels are not restricted to this wall thickness.*
- (5) The other super-pixels overlapped with the LA wall and PV but not selected as enhancement were considered as nonenhancement (overlapping ratio was set to  $\geq 20\%$ ).

By performing the five steps described above, we constructed a training dataset that contains super-pixels labeled either enhancement or nonenhancement within the LA wall and PV.

Instead of extracting texture or shape features of these labeled super-pixels, we computed the pixel intensity-based features to feed to the SVM classifier. This is because the size of our super-pixels is too small to catch enough information about texture and shape. In this study, we extracted 16 features for each super-pixel: minimum, maximum, mean, median, standard deviation, variance, mean of the absolute deviation, median absolute deviation, coefficient of variance, skewness, kurtosis, mode, central moments, range, interquartile range, and entropy. Feature selection was done using minimum redundancy and maximum relevance method.<sup>59</sup> In this study, we applied the mutual information quotient scheme.<sup>59</sup> The selected features will be presented in Section 3 and will be used for the further SVM-based classification procedure. The parameters of the SVM with a RBF kernel were optimized using cross-validation with a grid search scheme.<sup>60</sup>

## 2.E. Results evaluation and validation

### 2.E.1. Evaluation and validation of the MA-WHS

One experienced cardiologist (>5 yr of experience and specialized in cardiac MRI) manually segmented the

endocardial LA boundary and labeled the PV slice-by-slice in the LGE MRI images for all the patients. A second senior cardiologist (>25 yr of experience and specialized in cardiac MRI) confirmed the manual segmentation. The evaluation and validation of our MA-WHS has been done against this manual segmentation, which is assumed to be the ground truth. We used six metrics: Dice score, Jaccard index,<sup>61</sup> Precision, Negative Predictive Value (NPV), Hausdorff distance<sup>62</sup>, and Average Surface Distance (ASD) (defined in Table II for the delineated LA and PV ROIs). DICE, JACCARD, PRECISION, and Negative Predictive Value (NPV) measure the overlap (in percentage) between the manual and automatic segmentations. JACCARD is numerically more sensitive to mismatch when there is reasonably strong overlap than DICE or PRECISION. The higher the values of DICE, JACCARD, NPV, and PRECISION, the better the overall performance of the segmentation. HAUSDORFF and ASD measure the boundary distance (in mm) between two contours of segmentation. The lower the values of HAUSDORFF and ASD the better the agreement between manual delineation and fully automatic segmentation.

**2.E.2. Ground truth definition of the atrial scarring**

We formed the ground truth of the enhanced atrial scarring on the LGE MRI images using the following steps:

- (1) Steps (1)–(4) as listed in the “SVM Based Classification” section.
- (2) Once the enhanced super-pixels were extracted, they were combined to create a binary image for each slice, that is, 1 for enhanced super-pixels and 0 for unenhanced.
- (3) The binary image was overlaid on the original LGE MRI images and our cardiologist performed manual corrections to create the final boundaries (ground truth) of the enhanced atrial scarring. It should be stressed here that the cardiologist used the binary image only as an initial starting point for the manual segmentation in order to reduce an otherwise very lengthy process. While this could potentially introduce a bias toward the initial binary image, the cardiologist was free to make as many edits as needed and this bias would be expected to be small.

TABLE II. Summary of the quantitative evaluation methods.  $F_{Manual}$ : ground truth segmentation;  $F_{Auto}$ : automatic segmentation;  $|\bullet|$ : the number of pixels assigned to the segmentation;  $T$ : the total number of pixels;  $P_{Manual} = \{p_{m1}, \dots, p_{mn}\}$  and  $P_{Auto} = \{p_{a1}, \dots, p_{an}\}$ : two finite point sets of the two segmented contours (using the ground truth segmentation and automatic segmentation);  $\|\bullet\|$ :  $L_2$  norm; sup: supremum and inf: infimum. For the MA-WHS method we used all the six evaluation metrics and for the final atrial scarring segmentation we employed DICE, JACCARD, PRECISION, and NPV.

Evaluation metrics	Definition	MA-WHS	Atrial scarring segmentation
Dice score	$DICE = \frac{2 \times  F_{Manual} \cap F_{Auto} }{ F_{Manual}  +  F_{Auto} }$	•	•
Jaccard index	$JACCARD = \frac{ F_{Manual} \cap F_{Auto} }{ F_{Manual} \cup F_{Auto} }$	•	•
Precision	$PRECISION = \frac{ F_{Manual} \cap F_{Auto} }{ F_{Auto} }$	•	•
Negative Predictive Value	$NPV = \frac{T -  F_{Manual} \cup F_{Auto} }{T -  F_{Auto} }$	•	•
Hausdorff distance	$HAUSDORFF(P_{Manual}, P_{Auto}) = \max(d(P_{Manual}, P_{Auto}), d(P_{Auto}, P_{Manual}))$ <p>where <math>d(P_{Manual}, P_{Auto}) = \sup_{p_m \in P_{Manual}} \inf_{p_a \in P_{Auto}} \ p_m - p_a\ </math></p> $d(P_{Auto}, P_{Manual}) = \sup_{p_a \in P_{Auto}} \inf_{p_m \in P_{Manual}} \ p_m - p_a\ $	•	
Average Surface Distance	$ASD = \frac{1}{2} \left( \frac{\sum_{p_m \in P_{Manual}} \min_{p_a \in P_{Auto}} \ p_m - p_a\ }{\sum_{p_m \in P_{Manual}} 1} + \frac{\sum_{p_a \in P_{Auto}} \min_{p_m \in P_{Manual}} \ p_m - p_a\ }{\sum_{p_a \in P_{Auto}} 1} \right)$	•	

### 2.E.3. Intra- and interobserver variances of the manual atrial scarring segmentation

Two cardiologists performed the manual mouse clicks based ground truth construction procedures in 8 randomly selected patients (four pre- and four postablation cases; 235 2D images in total) in order to determine interobserver variance. In addition, one cardiologist performed the manual mouse clicks twice at two different time points (1 month in between) to estimate the intraobserver variance. The DICE metric was used to measure the intra- and interobserver variances of the ground truth construction.

### 2.E.4. Evaluation and validation of the fully automated atrial scarring segmentation

The SVM-based classification was evaluated by: (a) leave-one-patient-out cross-validation (LOO CV), which provides an unbiased predictor and is capable of creating sufficient training data for studies with small sample size;<sup>63</sup> (b) the cross-validated classification accuracy, sensitivity, specificity, and average area under the receiver operating characteristic (ROC) curve (AUC), and (c) the balanced error rate (BER).<sup>64</sup> We also applied 10-fold CV to evaluate the robustness of our method when there are fewer manual labeled training datasets. Lastly, we divided our data into (a) a training/CV dataset (25 patients) and (b) an independent testing dataset (12 patients). In this case, we optimized SVM parameters in the training/CV dataset and then validated our method using these parameters in the independent testing dataset. This validation demonstrates the robustness of our method on an “unseen” independent testing dataset after the SVM parameters have been optimized and fixed.

For the final atrial scarring segmentation, we also performed result evaluation using DICE, JACCARD, PRECISION, and NPV measurements (Table II). HAUSDORFF and ASD metrics were not applied because we have multiple discrete regions of enhanced atrial scarring for each LGE MRI volume.

In addition, the fibrosis extent percentage (FEP) was determined for ground truth and automatic segmentations and compared using two-sample Wilcoxon analysis. The FEP of the atrial scarring is an important imaging biomarker for predicting the outcome of the AF treatment. FEP is defined as the volume of scar tissue as a percentage of the atrial wall volume.<sup>15</sup>

### 2.E.5. Comparison study

In order to demonstrate the efficacy of our method, we also compared it with two standard methods published in previous studies:

- (a) Simple thresholding-based method (Thr).<sup>26</sup> The threshold value for each LGE MRI volume was chosen via empirical evaluation.

- (b) Conventional standard deviation (SD)<sup>10</sup> based method (2, 4, and 6 SDs were tested).

These methods were selected as they have minimum parameter tuning and could be most accurately reproduced, and also because they are the most frequently used.

In addition, we also compared our method with state-of-the-art unsupervised learning-based clustering and graph cuts-based methods<sup>5</sup> including k-means clustering-based method (KM), graph cuts with k-means clustering-based method (KM + GC), fuzzy c-means clustering-based method (FC), and graph cuts with fuzzy c-means clustering-based method (FCM + GC). In each case, we have reproduced the method as best possible, and where tuning parameters have not been clearly defined, we have used values based on those in the original methodology publications.

We compared the atrial scarring segmentation from each method against the ground truth using the LA and PV boundaries derived from our fully automatic MA-WHS segmentation. This was then repeated using the manually delineated LA and PV boundaries. The image intensities were normalized with respect to the mean and standard deviation of the intensities in the LA blood pool cavity,<sup>5</sup> which were extracted by a morphological erosion (5 mm) from the endocardial LA boundary.

For comparison studies, statistical significances were given by two-sample Wilcoxon rank-sum test.

## 3. RESULTS

### 3.A. Whole heart segmentation results

Figure 2(a) shows the comparison results of the Dice scores using the different label fusion schemes, for example, by majority vote (MV), by local weighted voting (LWV),<sup>37</sup> by joint label fusion (JLF),<sup>43</sup> by patch fusion one scale (PF), and by our proposed MSP. The MSP was significantly better than the other label fusion schemes ( $P < 0.05$ ). Figure 2(b) shows the quantitative results of this MA-WHS method compared to the ground truth.

### 3.B. Fully automated atrial scarring segmentation results

#### 3.B.1. Intra- and interobserver variances of the ground truth construction

Figure 3 demonstrates the intra- and interobserver variances of the manual atrial scarring delineation. Both intra- and interobserver agreement are very good (mean DICE scores ranging from 86% to 92% and from 83% to 91%, respectively) and confirm the suitability of the ground truth reconstructions for evaluation of the atrial scarring segmentation algorithms.



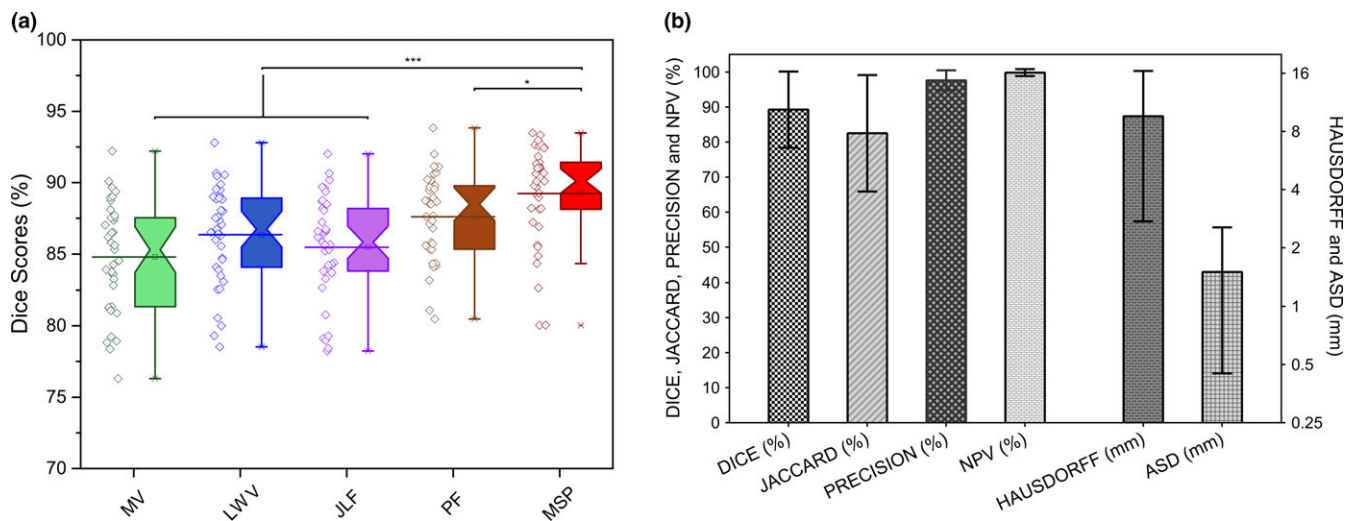


FIG. 2. (a) Comparison results (Dice scores of the WHS) of using different label fusion algorithms. (\*\*\*) =  $P < 0.05$  and (\*\*\*\*) =  $P < 0.0005$ ; statistical significances were given by two-sample Wilcoxon rank-sum test). Abbreviations: MV, majority vote; LWV, local weighted voting; JLF, joint label fusion; PF, patch fusion one scale; MSP, multiscale patch. (b) quantitative comparison of MA-WHS compared to manually determined ground truth. [Color figure can be viewed at wileyonlinelibrary.com]

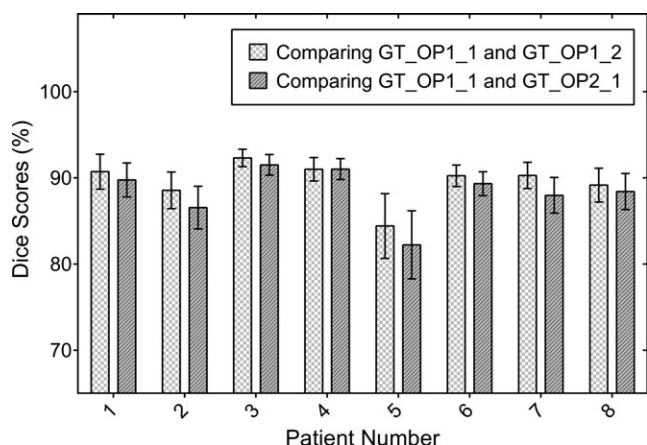


FIG. 3. Comparison of the manual delineations (ground truth) of the enhanced atrial scarring to demonstrate the inter- and intraobserver variances for eight randomly selected patient cases (235 2D slices in total). The mean and standard deviation (error bars) are shown. Abbreviations: GT\_OP1\_1: ground truth/operator 1/time point 1; GT\_OP1\_2: ground truth/operator 1/time point 2; GT\_OP2\_1: ground truth/operator 2/time point 1.

### 3.B.2. Evaluation and validation results of the fully automated atrial scarring segmentation

After minimum redundancy and maximum relevance based feature selection, 3 of 16 features (minimum, mean and standard deviation) were selected and used in building the SVM model. Table III tabulates the SVM classification results of distinguishing enhanced atrial scarring regions from nonenhanced tissues. Using LOO CV, we obtained 88% accuracy and 0.16 BER. ROC analysis shows an AUC of 0.91. In terms of the final segmentation accuracy, we achieved mean DICE of 79%. Compared to manual ground truth construction that took 25–50 min per patient case, the SVM-based prediction only took  $5.1 \pm 0.7$  s to segment one

patient case, while for a single loop of the LOO CV the training on 36 patients took  $64.2 \pm 4.6$  min (~1.8 min per patient case). All experiments were performed using a Windows 7 workstation with six-core 1.9 GHz Intel® Xeon® E5-2609v3/ 64 GB RAM.

Ten-fold cross-validation obtained similar accuracy. For the validation on the separated independent testing datasets, the fixed SVM model was blindly built. Compared to LOO CV on 37 datasets we achieved similar accuracy (86%) and sensitivity (92%), but lower specificity (64%) and mean DICE (71%). This validation using separate testing datasets showed that our method can still perform well while fixing the classification model and using it to segment the new input data.

Using our fully automatic pipeline, the measured native (preablation) fibrosis associated with AF was  $26.9 \pm 11.2\%$  compared to  $32.8 \pm 6.4\%$  for the postablation cases [Fig. 4(a)]. There was no significant difference found for the fibrosis extent derived using the ground truth segmentation of the atrial scarring ( $23.4 \pm 7.3\%$  for the preablation cases and  $30.9 \pm 6.1\%$  for the postablation cases). The high fibrosis extent measured by both techniques in the preablation patients reflects the extensive native fibrosis due to atrial remodeling in this patient cohort with long-standing persistent AF. Both our fully automatic pipeline and the ground truth segmentation found significant differences in fibrosis extent between pre- and postablation cases [Fig. 4(a)]. Bland–Altman analyses are shown in Figs. 4(b) and 4(c).

Figure 5 shows the results of the comparison study in preablation [Figs. 5(a) and 5(c)] and postablation [Figs. 5(b) and 5(d)] cases with our method (red bars) working equally well in both (median DICE score 80% for the postablation cases vs. median DICE score 76% for the pre-ablation cases,  $P = 0.087$ ). Overall, the atrial scarring segmentation results obtained using our method outperformed the simple

TABLE III. Quantitative evaluation of the atrial scarring segmentation using LOO CV, 10-fold CV, and training/CV with separate testing. The SVM based classification was evaluated using accuracy, sensitivity, specificity, BER, and AUC. The final segmentation was evaluated against the ground truth using Precision, NPV, Jaccard index and Dice score. Abbreviations: LOO, leave-one-patient-out; CV, cross-validation; BER, balanced error rate; AUC, area under curve; NPV, negative predictive value.

Validation method		Accuracy (%)	Sensitivity (%)	Specificity (%)	BER	AUC	Precision (%)	NPV (%)	Jaccard Index (%)	Dice Score (%)
LOO CV (37 patients)		88	90	79	0.16	0.91	81 ± 9	99 ± 1	65 ± 6	79 ± 5
10-fold CV (37 patients)		88	96	62	0.21	0.91	86 ± 4	99 ± 2	56 ± 3	72 ± 2
Training/CV + separate testing	LOO CV (25 patients)	87	89	79	0.16	0.91	80 ± 10	99 ± 1	66 ± 6	79 ± 5
	Separate testing (12 patients)	86	92	64	0.22	0.88	77 ± 7	99 ± 1	56 ± 7	71 ± 7

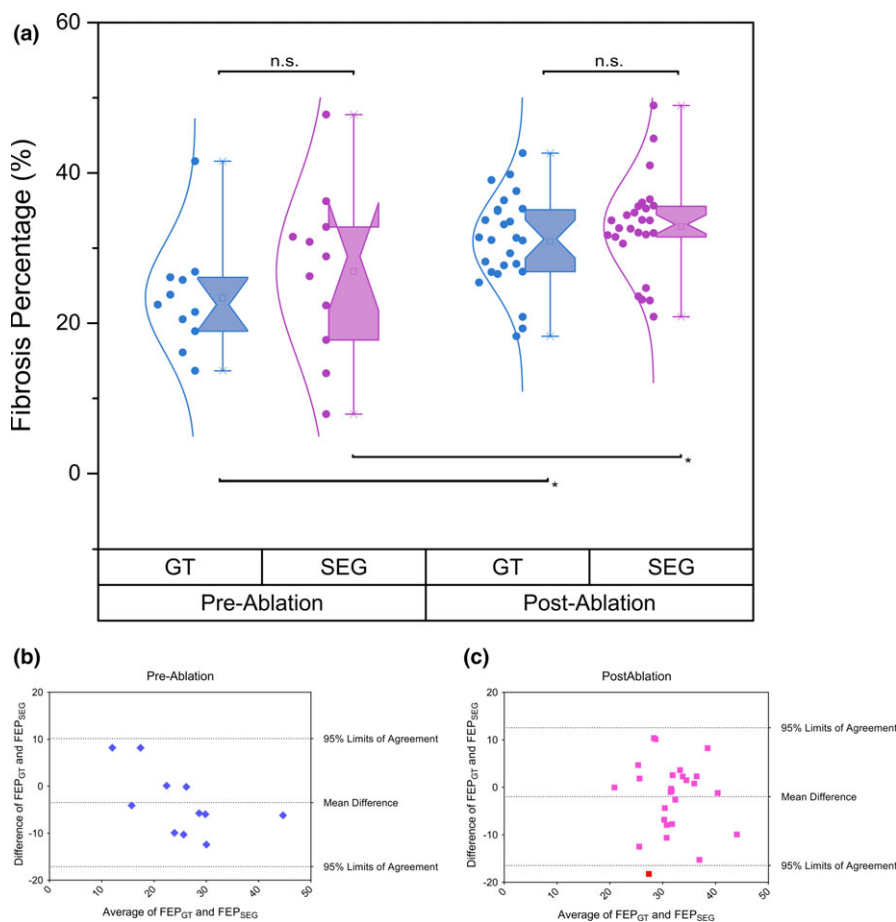


FIG. 4. (a) The percentage of fibrosis extent calculated using the ground truth segmentation (GT in blue boxplots) and using our fully automatic segmentation pipeline (SEG in purple boxplots) for both the preablation cases and the postablation cases, respectively. (“\*” =  $P < 0.05$  and “n.s.” means no significant difference between two groups; statistical significances were given by two-sample Wilcoxon rank-sum test). Abbreviations: GT, ground truth segmentation; SEG, segmentation using our fully automatic pipeline. (b)–(c) Bland–Altman analysis of the measurements of fibrosis extent derived by using the ground truth segmentation (FEP<sub>GT</sub>) and our fully automatic segmentation pipeline (FEP<sub>SEG</sub>). Abbreviations: GT, ground truth segmentation; SEG, segmentation using our fully automatic pipeline; FEP, fibrosis extent percentage. [Color figure can be viewed at wileyonlinelibrary.com]

thresholding and conventional standard deviation methods significantly [Figs. 5(a) and 5(b)]. While, as expected, DICE scores for the state-of-the-art unsupervised learning based clustering and graph-cuts based methods were better than thresholding and standard deviation based methods [Figs. 5(c) and 5(d)], our method again showed superior results.

Figure 6 shows our final atrial scarring segmentation results compared to the ground truth in two preablation and two postablation patients. In each case, the figure shows a single slice of a 3D LGE dataset (typically 64 slices) while the FEP values quoted are for the full 3D study. The segmentation results have been derived from the LOO CV (i.e., training on 36 datasets and making prediction on the one

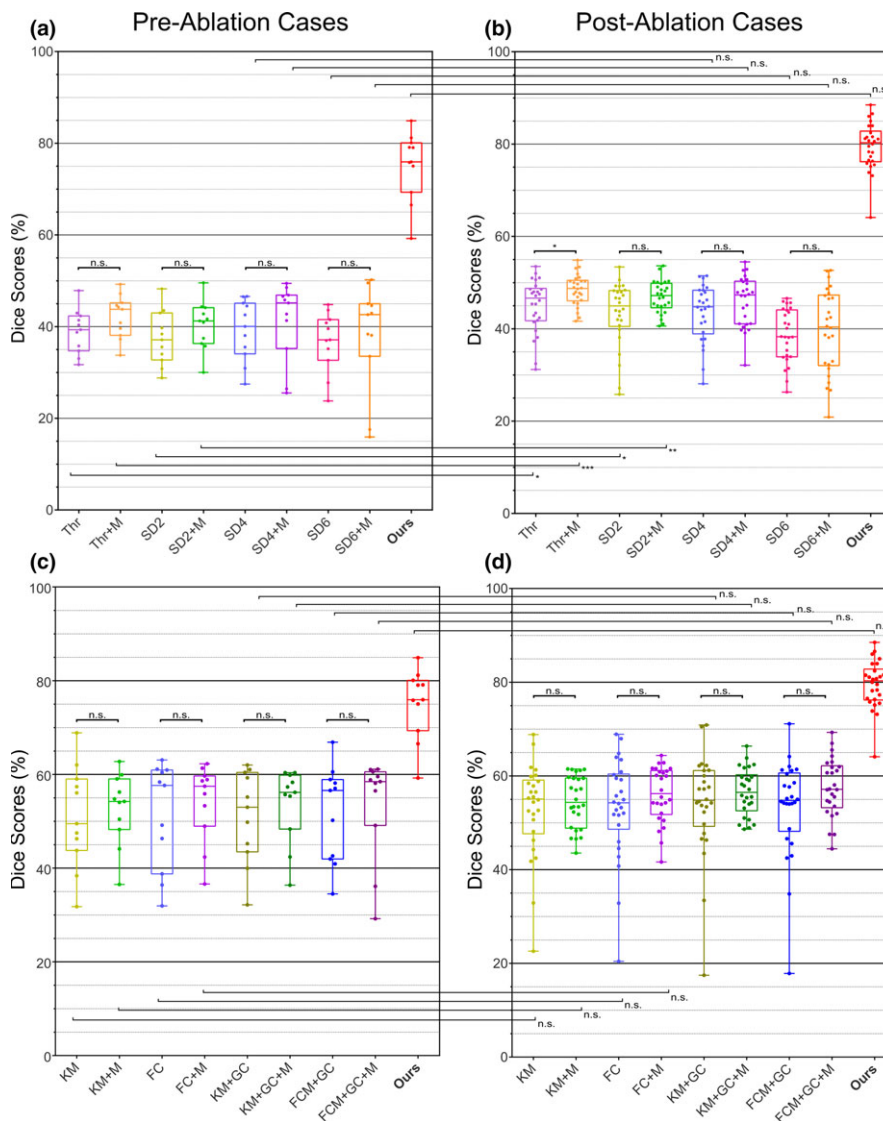


FIG. 5. Comparison results with conventional atrial scarring segmentation methods using DICE for (a) and (c) preablation and (b) and (d) postablation cases. (“\*” =  $P < 0.05$ , “\*\*” =  $P < 0.005$ , “\*\*\*” =  $P < 0.0005$ , and “n.s.” means no significant difference between two groups; statistical significances were given by two-sample Wilcoxon rank-sum test). Abbreviations: Thr, simple thresholding-based method with MA-WHS-derived LA + PV; SD(x), conventional standard deviation method (x = 2, 4 and 6 SDs were tested) with MA-WHS-derived LA + PV; KM, k-means clustering-based method with MA-WHS-derived LA + PV; FC, fuzzy C-means clustering-based method with MA-WHS-derived LA + PV; KM + GC, graph cuts and k-means clustering with MA-WHS-derived LA + PV; FCM + GC, graph cuts and fuzzy C-means clustering with MA-WHS-derived LA + PV; (x) + M, different methods with manual delineated LA + PV; MA-WHS, multias atlas whole heart segmentation; LA + PV, left atrium and pulmonary veins. [Color figure can be viewed at wileyonlinelibrary.com]

dataset that has been left). For the first preablation case [Figs. 6(a)–6(c)], the automatic segmentation slightly underestimates the enhancement (FEP 9.2% compared to 12.8%). Segmentation of the second preablation case [Figs. 6(d)–6(f)] shows clear accordance compared to the ground truth despite some slight overestimation near the right inferior pulmonary vein. Segmentation of both postablation cases exhibit good agreement with the ground truth [Figs. 6(i) vs. 6(h) and 6(l) vs. 6(k)].

#### 4. DISCUSSION

We have developed a novel fully automatic segmentation pipeline to detect atrial scarring in LGE MRI images and

validated it against manual ground truth segmentation by experienced cardiologists.

Accurate knowledge of native and previous ablation scarring using LGE MRI may be used to stratify patients and guide ablation treatment, ultimately reducing the need for repeat procedures.<sup>5</sup> Gadolinium-based contrast agent is contraindicated in patients with severe renal impairment due to an increased risk of developing nephrogenic systemic fibrosis, a rare but serious disease. In addition, there have been recent reports describing the retention of gadolinium in the body for up to several months after administration, even in patients with normal kidney function. This has led to the Federal Drugs Administration (FDA) issuing a safety communication and while no direct association has been found

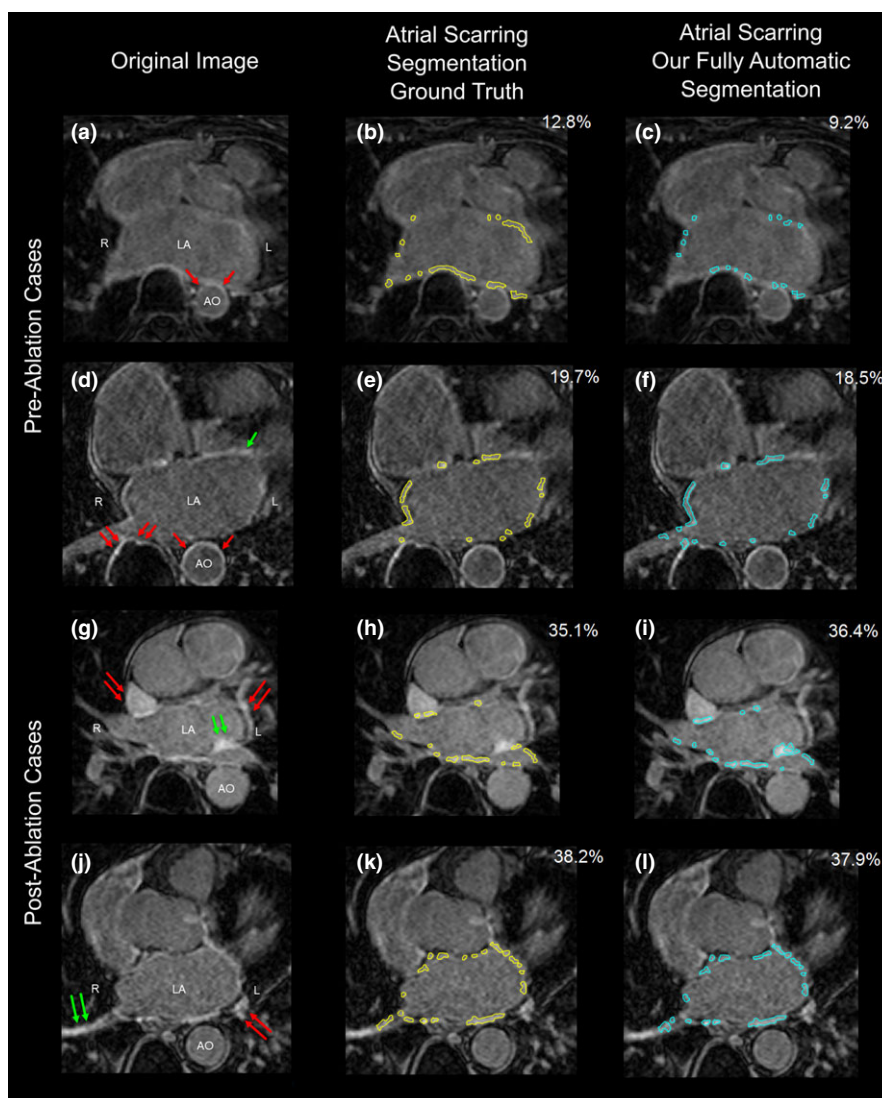


FIG. 6. Final atrial scarring segmentation results of two preablation cases (a–c) and (d–f) and two postablation cases (g–i) and (j–l). (a), (d), (g), and (j) Original LGE MRI images; (b), (e), (h), and (k) Ground truth of the atrial scarring segmentation; (c), (f), (i), and (l) Results of our fully automatic atrial scarring segmentation. Single red arrows in (a) and (d) show the enhancement of the AO wall. Double red arrows in (d), (g), and (j) show enhanced regions outside the heart, in other substructures inside the heart or in surrounding fat tissue. Single green arrow in (d) shows the enhanced artifacts of the mitral valve. Double green arrows in (g) show some enhancement that, in retrospect, our cardiology experts agreed should have been included in the ground truth labeling procedure, but was found using our fully automatic segmentation. And double green arrows in (j) show the enhancement due to the navigator beam and blood flow. Abbreviations: LA, left atrium; AO, aorta; L, left; R, right. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

between these deposits and adverse health effects in patients with normal kidney function, it has announced that a new class warning and safety measures will be required to appear in the labeling for gadolinium contrast agents used in MRI. Consideration should be given to the use of macrocyclic gadolinium agents, such as gadobutrol (used in this study), rather than linear agents as they have reduced and shorter retention properties. The FDA concludes that the benefit of all approved gadolinium contrast agents continues to outweigh potential risks and that healthcare professionals should not avoid or defer necessary examinations.<sup>65</sup>

Segmentation of the atrial scarring from LGE MRI images is very challenging. This is not only because the atrial scarring is difficult to distinguish in the thin LA wall but also

because the image quality can be poor due to motion artifacts, noise contamination and contrast agent wash-out during the long acquisition. Moreover, the enhancement from the surrounding tissues and enhanced blood flow can result in increased false positives. However, most of these confounding enhancement regions can be distinguished subject to accurate heart anatomy delineation using our MA-WHS [Figs. 6(c), 6(f), 6(i), and 6(l)].

The minimum redundancy and maximum relevance method has selected three simple but effective features for our further SVM classification on SLIC segmented super-pixels, that is, the mean, the standard deviation, and the min of the super-pixels. The feature “mean” corresponded to a simple thresholding on the super-pixel intensity values and



the feature “standard deviation” quantified local intensity variations in the scar and healthy regions. The feature “min” was selected as a strong discriminator due to the fact that in the enhanced atrial scarring regions the “min” intensity values are much higher than the “min” intensity values of the normal regions. However, the feature “max” was not selected as a discriminator as for some labeled normal regions, there may be relatively high pixel intensities that could be false positives.

There are limitations of the current work. The fast and irregular heart rate in patients prior to ablation resulted in only 11 (65%) preablation studies having good enough quality to be included in this study. Together with 26 (92%) postablation studies, our total number of datasets was limited to 37. The low number of preablation studies included is an indication of how difficult it is to scan patients in AF. However, developments in MRI acquisition strategies (beat-to-beat changes in inversion time and intelligent arrhythmia rejection schemes) are likely to improve this. To tackle the problem of having limited patient data, LOO CV was performed to achieve an unbiased predictor for limited datasets.<sup>66,67</sup>

As in other studies,<sup>5</sup> we have validated our technique against a manually segmented ground truth. In our case, to reduce a very lengthy process, the starting point for this was an automatically derived binary image which the cardiologist was free to edit as much as possible. While this could introduce some bias into the ground truth, this is expected to be small and the methodology had both low inter- and intraoperator variance. However, even with experienced cardiologists and consensus agreement, this is subjective, and regions of enhancement can be missed [Fig. 6(g)], but they can still be found using our fully automatic segmentation pipeline. EAM may provide a more objective ground truth but is not straightforward due to issues relating to sampling point density and distribution, contact force which affects the measured voltage, voltage thresholds for delineating scar tissue, and registration and deformation of the mapped space in the EP lab to the LGE MRI study.

In this study, we compared with the simple thresholding and conventional standard deviation based methods and also a number of advanced unsupervised learning based clustering and graph cuts-based methods.<sup>5</sup> Although superior performance has been achieved using our propose method, it is of note that our patient cohort is different from that in which these comparison algorithms were optimized and tested. When compared to manual segmentation (ground truth) in postablation scans, these standard techniques gave median DICE of 38%–48% while our fully automatic technique achieved a median DICE of 80%. The results that we obtained here with the standard techniques are similar to those reported with these same techniques in the benchmarking study described in Ref. [5] while the latter score is similar to the best-performing methods reported in that same study. Moreover, performance of the state-of-the-art unsupervised learning based methods were better than thresholding and standard deviation-based methods [Figs. 5(c) and 5(d)], but our method again showed

superior results that might be attributed to the fact that our proposed method is supervised, and therefore have obtained more useful information from the manually labeled data. Of note is that in the benchmarking study, the variances of all of the techniques tested are large while in our manuscript, the results are more consistent with a relatively small variance (boxplot in red as seen in Fig. 5). This may be due to our patient cohort being more tightly defined while in the previous study, datasets were analyzed from patients at multiple institutions using a variety of imaging protocols.

## 5. CONCLUSIONS

To the best of our knowledge, this is the first study that developed a fully automatic segmentation pipeline for atrial scarring segmentation with quantitative validation for LGE MRI scans. The proposed pipeline has demonstrated an effective and efficient way to objectively segment and assess the atrial scarring. Our validation results have shown that both our MA-WHS and super-pixel classification-based atrial scarring segmentation have obtained satisfactory accuracy. The current study was performed using real clinical data, and we can envisage an integration of our pipeline to clinical routines. In so doing, a patient-specific LA and PV geometry model and an objective atrial scarring segmentation can be obtained rapidly for individual AF patients without manual processing.

## ACKNOWLEDGMENTS

This study was funded by the British Heart Foundation Project Grant (Project Number: PG/16/78/32402), the NIHR Cardiovascular Biomedical Research Unit, Royal Brompton Hospital & Harefield NHS Foundation Trust and Imperial College London, and the Science and Technology Commission of Shanghai Municipality (17JC1401600), and the NSFC-RS fund (81511130090). Data were obtained during the NIHR Efficacy and Mechanism Evaluation Programme (Project Number: 12/127/127).

## CONFLICT OF INTEREST

The authors have no conflicts to disclose.

\*Joint senior authors.

<sup>a)</sup>Authors to whom correspondence should be addressed. Electronic mails: g.yang@imperial.ac.uk; zxh@fudan.edu.cn; Telephone: +(0044) 020 7352 8121; Fax: (0044) 020 7351 8699.

## REFERENCES

1. Allessie M. Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovasc Res.* 2002;54:230–246.
2. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary. *J Am Coll Cardiol.* 2014;64:2246–2280.
3. Ouyang F, Tilz R, Chun J, et al. Long-term results of catheter ablation in paroxysmal atrial fibrillation: lessons from a 5-year follow-up. *Circulation.* 2010;122:2368–2377.

4. Weerasooriya R, Khairy P, Litalien J, et al. Catheter ablation for atrial fibrillation: are results maintained at 5 years of follow-up? *J Am Coll Cardiol*. 2011;57:160–166.
5. Karim R, Housden RJ, Balasubramaniam M, et al. Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge. *J Cardiovasc Magn Reson*. 2013;15:105–122.
6. Calkins H, Kuck KH, Cappato R, Heart Rhythm Society Task Force on Catheter and Surgical Ablation of Atrial Fibrillation, et al. 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. *Heart Rhythm*. 2012;9:632–696.e21.
7. Zhong H, Lacomis JM, Schwartzman D. On the accuracy of CartoMerge for guiding posterior left atrial ablation in man. *Heart Rhythm*. May 2007;4:595–602.
8. Schmidt EJ, Mallozzi RP, Thiagalingam A, et al. Electroanatomic mapping and radiofrequency ablation of porcine left atria and atrioventricular nodes using magnetic resonance catheter tracking. *Circ Arrhythm Electrophysiol*. 2009;2:695–704.
9. Simonetti OP, Kim RJ, Fieno DS, et al. An improved MR imaging technique for the visualization of myocardial infarction. *Radiology*. 2001;218:215–223.
10. Oakes RS, Badger TJ, Kholmovski EG, et al. Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. *Circulation*. 2009;119:1758–1767.
11. Marrouche NF, Wilber D, Hindricks G, et al. Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: the DECAAF study. *JAMA*. 2014;311:498–506.
12. Akoum N, Daccarett M, McGann C, et al. Atrial fibrosis helps select the appropriate patient and strategy in catheter ablation of atrial fibrillation: a DE-MRI guided approach. *J Cardiovasc Electrophysiol*. 2011;22:16–22.
13. Vergara GR, Marrouche NF. Tailored management of atrial fibrillation using a LGE-MRI based model: from the clinic to the electrophysiology laboratory. *J Cardiovasc Electrophysiol*. 2011;22:481–487.
14. Malcolme-Lawes LC, Juli C, Karim R, et al. Automated analysis of atrial late gadolinium enhancement imaging that correlates with endocardial voltage and clinical outcomes: a 2-center study. *Heart Rhythm*. 2013;10:1184–1189.
15. Ravanelli D, Dal Piaz EC, Centonze M, et al. A novel skeleton based quantification and 3-D volumetric visualization of left atrium fibrosis using late gadolinium enhancement magnetic resonance imaging. *IEEE Trans Med Imaging*. 2014;33:566–576.
16. McGann C, Akoum N, Patel A, et al. Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI. *Circ Arrhythmia Electrophysiol*. 2014;7:23–30.
17. Jadidi AS, Cochet H, Shah AJ, et al. Inverse relationship between fractionated electrograms and atrial fibrosis in persistent atrial fibrillation: combined magnetic resonance imaging and high-density mapping. *J Am Coll Cardiol*. 2013;62:802–812.
18. Reddy VY, Schmidt EJ, Holmvang G, Fung M. Arrhythmia recurrence after atrial fibrillation ablation: can magnetic resonance imaging identify gaps in atrial ablation lines? *J Cardiovasc Electrophysiol*. 2008;19:434–437.
19. Badger TJ, Daccarett M, Akoum NW, et al. Evaluation of left atrial lesions after initial and repeat atrial fibrillation ablation; lessons learned from delayed-enhancement MRI in repeat ablation procedures. *Circ Arrhythmia Electrophysiol*. 2010;3:249–259.
20. Kapa S, Desjardins B, Callans DJ, Marchlinski FE, Dixit S. Contact electroanatomic mapping derived voltage criteria for characterizing left atrial scar in patients undergoing ablation for atrial fibrillation. *J Cardiovasc Electrophysiol*. 2014;25:1044–1052.
21. Bisbal F, Guiu E, Cabanas-Grandío P, et al. CMR-guided approach to localize and ablate gaps in repeat AF ablation procedure. *JACC Cardiovasc Imaging*. 2014;7:653–663.
22. Harrison JL, Jensen HK, Peel SA, et al. Cardiac magnetic resonance and electroanatomical mapping of acute and chronic atrial ablation injury: a histological validation study. *Eur Heart J*. 2014;35:1486–1495.
23. Knowles BR, Caulfield D, Cooklin M, et al. 3-D visualization of acute RF ablation lesions using MRI for the simultaneous determination of the patterns of necrosis and edema. *IEEE Trans Biomed Eng*. 2010;57:1467–1475.
24. Tao Q, Ipek EG, Shahzad R, Berendsen FF, Nazarian S, van der Geest RJ. Fully automatic segmentation of left atrium and pulmonary veins in late gadolinium-enhanced MRI: towards objective atrial scar assessment. *J Magn Reson Imaging*. 2016;44:346–354.
25. Karim R, Arujuna A, Housden RJ, et al. A method to standardize quantification of left atrial scar from delayed-enhancement MR Images. *Transl Eng Heal Med IEEE J*. 2014;2:1–15.
26. Peters DC, Wylie JV, Hauser TH, et al. Recurrence of atrial fibrillation correlates with the extent of post-procedural late gadolinium enhancement. A pilot study. *JACC Cardiovasc Imaging*. 2009;2:308–316.
27. Peters DC, Wylie JV, Hauser TH, et al. Detection of pulmonary vein and left atrial scar after catheter ablation with three-dimensional navigator-gated delayed enhancement MR imaging: initial experience. *Radiology*. 2007;243:690–695.
28. Keegan J, Jhooti P, Babu-Narayan SV, Drivas P, Ernst S, Firmin DN. Improved respiratory efficiency of 3D late gadolinium enhancement imaging using the continuously adaptive windowing strategy (CLAWS). *Magn Reson Med*. 2014;71:1064–1074.
29. Keegan J, Drivas P, Firmin DN. Navigator artifact reduction in three-dimensional late gadolinium enhancement imaging of the atria. *Magn Reson Med*. 2013;75:779–785.
30. Moghari MH, Peters DC, Smink J, et al. Pulmonary vein inflow artifact reduction for free-breathing left atrium late gadolinium enhancement. *Magn Reson Med*. 2011;66:180–186.
31. Krishnam MS, Tomasian A, Deshpande V, et al. Noncontrast 3D steady-state free-precession magnetic resonance angiography of the whole chest using nonselective radiofrequency excitation over a large field of view: comparison with single-phase 3D contrast-enhanced magnetic resonance angiography. *Invest Radiol*. 2008;43:411–420.
32. Zhuang X, Bai W, Song J, et al. Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection. *Med Phys*. 2015;42:3822–3833.
33. Zhuang X, Shen J. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Med Image Anal*. 2016;31:77–87.
34. Tobon-Gomez C, Geers A, Peters J, et al. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE Trans Med Imaging*. 2015;34:1460–1473.
35. Zhuang X, Rhode KS, Razavi RS, Hawkes DJ, Ourselin S. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans Med Imaging*. 2010;29:1612–1625.
36. Zhuang X, Arridge S, Hawkes DJ, Ourselin S. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Trans Med Imaging*. 2011;30:1819–1828.
37. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE Trans Med Imaging*. 2009;28:1266–1277.
38. Bai W, Shi W, Ledig C, Rueckert D. Multi-atlas segmentation with augmented features for cardiac MR images. *Med Image Anal*. 2015;19:98–109.
39. Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL. Non-local patch-based label fusion for hippocampus segmentation. *Med Image Comput Comput-Assist Interv*. 2010;6361:129–136.
40. Eskildsen SF, Coupé P, Fonov V, et al. BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage*. 2012;59:2362–2373.
41. Hu S, Coupé P, Pruessner JC, Collins DL. Nonlocal regularization for active appearance model: application to medial temporal lobe segmentation. *Hum Brain Mapp*. 2014;35:377–395.
42. Sabuncu MR, Yeo BTT, Van Leemput K, Fischl B, Golland P. A generative model for image segmentation based on label fusion. *IEEE Trans Med Imaging*. 2010;29:1714–1729.
43. Wang H, Suh JW, Das SR, Pluta JB, Craige C, Yushkevich PA. Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell*. 2013;35:611–623.
44. Wu G, Kim M, Sanroma G, Wang Q, Munsell BC, Shen D. Alzheimer's Disease Neuroimaging Initiative. Hierarchical multi-atlas label fusion

- with multi-scale feature representation and label-specific patch partition. *NeuroImage*. 2015;106:34–46.
45. Lindeberg T. Feature detection with automatic scale selection. *Int J Comput Vis*. 1998;30:79–116.
  46. Lindeberg T. Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. *J Math Imaging Vis*. 2011;40:36–81.
  47. Holden M, Griffin LD, Saeed N, Hill DLG. Multi-channel Mutual Information Using Scale Space. In: *Medical Image Computing and Computer-Assisted Intervention*. France: Springer; 2004, pp. 797–804.
  48. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*. 2004;60:91–110.
  49. Leutenegger S, Chli M, Siegwart RY. BRISK: Binary Robust invariant scalable keypoints. In: 2011 International Conference on Computer Vision; 2011, pp. 2548–2555.
  50. Wu Y, Liu G, Huang M, et al. Prostate segmentation based on variant scale patch and local independent projection. *IEEE Trans Med Imaging*. 2014;33:1290–1303.
  51. Thévenaz P, Unser M. Optimization of mutual information for multiresolution image registration. *IEEE Trans Image Process*. 2000;9:2083–2099.
  52. Achanta R, Shaji A, Smith K, Lucchi A. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell*. 2012;34:2274–2281.
  53. Ye X, Beddoe G, Slabaugh G. Automatic graph cut segmentation of lesions in CT using mean shift superpixels. *Int J Biomed Imaging*. 2010;2010:983963.
  54. Soltaninejad M, Yang G, Lambrou T, et al. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *Int J Comput Assist Radiol Surg*. 2016;12:183–203.
  55. Zhao Y, Zheng Y, Liu Y, et al. Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy. *IEEE Trans Med Imaging*. 2017;36:51–63.
  56. Lemola K, Sneider M, Desjardins B, et al. Computed tomographic analysis of the anatomy of the left atrium and the esophagus: implications for left atrial catheter ablation. *Circulation*. 2004;110:3655–3660.
  57. Akçakaya M, Rayatzadeh H, Basha T, et al. Accelerated late gadolinium enhanced cardiac mr imaging with isotropic spatial resolution using compressed sensing: initial experience. *Radiology*. 2012;264:691–699.
  58. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: First. Cambridge University Press; 2000.
  59. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Anal Mach Intell IEEE Trans*. 2005;27:1226–1238.
  60. Hsu C, Chang C, Lin C. A practical guide to support vector classification; 2003.
  61. Zhuang X. Challenges and methodologies of fully automatic whole heart segmentation: a review. *J Healthc Eng*. 2013;4:371–408.
  62. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 1993;15:850–863.
  63. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc*. 1983;78:316.
  64. Chen Y, Lin C. Combining SVMs with various feature selection strategies. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA, eds. *Feature Extraction: Foundations and Applications*. Berlin, Heidelberg: Springer; 2006:315–324.
  65. FDA. FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings; 2017.
  66. Yang G, Jones TL, Howe FA, Barrick TR. Morphometric model for discrimination between glioblastoma multiforme and solitary metastasis using three-dimensional shape analysis. *Magn Reson Med*. 2015;75:2505–2516.
  67. Yang G, Jones TL, Barrick TR, Howe FA. Discrimination between glioblastoma multiforme and solitary metastasis using morphological features derived from the p;q tensor decomposition of diffusion tensor imaging. *NMR Biomed*. 2014;27:1103–1111.
  68. Perry D, Morris A, Burgon N, McGann C, MacLeod R, Cates J. Automatic classification of scar tissue in late gadolinium enhancement cardiac MRI for the assessment of left-atrial wall injury after radiofrequency ablation. In *SPIE Medical Imaging 2012: Computer-Aided Diagnosis*. Vol. 8315. Bellingham: International Society for Optics and Photonics; 2012: p. 83151D.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**Appendix S1.** Supporting material.