Automated Myocardial T2 and Extracellular Volume Quantification in Cardiac MRI Using Transfer Learning–based Myocardium Segmentation

Yanjie Zhu, PhD* • Ahmed S. Fahmy, PhD* • Chong Duan, PhD • Shiro Nakamori, MD • Reza Nezafat, PhD

Purpose: To assess the performance of an automated myocardial T2 and extracellular volume (ECV) quantification method using transfer learning of a fully convolutional neural network (CNN) pretrained to segment the myocardium on T1 mapping images.

Materials and Methods: A single CNN previously trained and tested using 11,550 manually segmented native T1-weighted images was used to segment the myocardium for automated myocardial T2 and ECV quantification. Reference measurements from 1,525 manually processed T2 maps and 1,525 ECV maps (from 305 patients) were used to evaluate the performance of the pretrained network. Correlation coefficient (R) and Bland-Altman analysis were used to assess agreement between automated and reference values on per-patient, per-slice, and per-segment analyses. Furthermore, transfer learning effectiveness in the CNN was evaluated by comparing its performance to four CNNs trained using manually segmented T2-weighted and postcontrast T1-weighted images and initialized using random-weights or weights of the pretrained CNN.

Results: T2 and ECV measurements using the pretrained CNN strongly correlated with reference values in per-patient (T2: R = 0.88, 95% confidence interval [CI]: 0.85, 0.91; ECV: R = 0.91, 95% CI: 0.89, 0.93), per-slice (T2: R = 0.83, 95% CI: 0.81, 0.85; ECV: R = 0.84, 95% CI: 0.82, 0.86), and per-segment (T2: R = 0.75, 95% CI: 0.74, 0.77; ECV: R = 0.76, 95% CI: 0.75, 0.77) analyses. In Bland-Altman analysis, the automatic and reference values were in good agreement in per-patient (T2: 0.3 msec ± 2.9; ECV: −0.3% ± 1.7), per-slice (T2: 0.1 msec ± 4.6; ECV: −0.3% ± 2.5), and per-segment (T2: 0.0 msec ± 6.5; ECV: −0.4% ± 3.5) analyses. The performance of the pretrained network was comparable to networks refined or trained from scratch using additional manually segmented images.

Conclusion: Transfer learning extends the utility of pretrained CNN-based automated native T1 mapping analysis to T2 and ECV mapping without compromising performance.

Supplemental material is available for this article.

©RSNA, 2020
Recently, a deep learning–based analysis platform was developed for automating analysis of native myocardial T1 mapping (26). T1 mapping data acquired by slice-interleaved myocardial T1 mapping sequence (STONE) (7) were used for training and validation. While similar network architectures could potentially automate other tissue parameters, such as ECV or T2 mapping, doing so would require a newly labeled dataset and a dedicated neural network for each individual parameter and sequence, which is not clinically feasible.

In this study, we sought to evaluate the performance of a pre-trained deep neural network using native T1 mapping datasets to automate T2 and ECV measurements. Three neural networks with similar architectures were employed to analyze each parameter: (a) the native T1 pretrained network; (b) a refined native T1 pretrained network (trained using a small dataset of T2-weighted and postcontrast T1-weighted images); and (c) a new network trained from scratch using a small dataset of T2-weighted and postcontrast T1-weighted images. Network performance was evaluated with respect to corresponding reference measurements computed by manual analysis.

Materials and Methods

Processing Pipeline

The proposed analysis pipeline (Fig 1) begins with automatic segmentation of the myocardium in each of the input of T1- or T2-weighted images. A CNN with a U-Net architecture (27) was used to segment the input images and produce a binary image containing a single object (ie, myocardium). If more than one object is segmented, only the largest object is maintained. To ensure that the segmented object is indeed the myocardium, two conditions are imposed on two automatically obtained. To ensure that the segmented object is indeed the myocardium, two conditions are imposed on two automatically acquired cardiac native T1 mapping images to be used to automate myocardial T2 and extracellular volume (ECV) quantification without a need for retraining with separate T2 or ECV mapping images.

An automated myocardial tissue mapping platform using deep learning–based image segmentation has potential to facilitate the clinical utility and standardization of quantitative myocardial mapping.

**Key Points**

- Convolutional neural networks trained to segment the myocardium using cardiac native T1 mapping images can be used to automate myocardial T2 and extracellular volume (ECV) quantification without a need for retraining with separate T2 or ECV mapping images.
- An automated myocardial tissue mapping platform using deep learning–based image segmentation has potential to facilitate the clinical utility and standardization of quantitative myocardial mapping.

**Abbreviations**

CI = confidence interval, CNN = convolutional neural network, ECV = extracellular volume, STONE = slice-interleaved myocardial T1 mapping sequence

**Summary**

A single convolutional neural network, trained for myocardium segmentation on T1 maps, can be used to automate myocardial T2 and extracellular volume maps with comparable accuracy to that of manual assessment.

**Image Acquisition**

Patients known to have or suspected of having cardiovascular disease referred for clinical cardiac MRI were prospectively recruited from 2016 to 2017. The institutional review board approved the study, written informed consent was obtained from each patient prior to examination, and patient data handling was Health Insurance Portability and Accountability Act compliant. A total of 305 patients (203 male patients and 102 female patients; mean age: 55 years ± 15 [standard deviation]) were recruited for myocardial T2 and ECV mapping (see acquisition parameters in Appendix E1 [supplement]).

**Training and validation images.**—T1- and T2-weighted images from 305 patients were randomly divided into (a) a network training subset (21 patients; 945 T2-weighted images and 1155 postcontrast T1-weighted images); (b) a validation subset (284 patients; 12780 T2-weighted images and 15620 postcontrast T1-weighted images); and (c) a network testing subset randomly selected from the validation subset (30 patients; 1650 postcontrast T1-weighted images and 1350 T2-weighted images). Network training and validation subsets were mutually exclusive. All images were standardized to 256 × 256 pixels using bicubic interpolation (28). Network training and...
steps were performed using the automated analysis platform. We evaluated agreement between automated T2 and ECV values and corresponding reference values using linear regression.
and Pearson correlation and Bland-Altman analyses. Each network was evaluated according to per-patient and per-slice success rate, defined as percentage of successfully generated maps relative to the total number of patients and total number of slices, respectively. Mapping of a given slice was considered successful if the network successfully segmented the myocardium on at least six T2-weighted images (for T2 mapping), or on at least eight native and eight postcontrast T1-weighted images (for ECV mapping). Mapping of a given patient was considered successful if at least one apical, one midventricular, and one basal slice were successfully computed. The Wilcoxon signed-rank test was conducted to compare automatic and reference measurements. Statistical significance was set to \( P < .05 \). Additionally, we calculated the sensitivity and specificity of CNN\(_{\text{nd}}\)\(_{\text{T1}}\) for identifying patients with myocardium mapping parameters above a cutoff threshold of 58 msec for T2 mapping and above 29% for ECV mapping. The threshold values were arbitrarily set to mean + 2 standard deviations of the parameter values reported in normal populations (29,30). Dice similarity coefficient was used to assess the CNN segmentation accuracy by measuring the overlap between the CNN-segmented and the manually segmented myocardium.

### Results

The pretrained CNN\(_{\text{nd}}\)\(_{\text{T1}}\) network delineated the epicardium and endocardium for different slices, inversion times, and TE\(_{\text{T2prep}}\) times (Fig 2). Representative T2 and ECV maps reconstructed using CNN\(_{\text{nd}}\)\(_{\text{T1}}\) are shown (Fig 3). In Exp#1, the success rate for reconstructing T2 maps was 276 of 284 (97%) (per-patient) and 1329 of 1420 (94%) (per-slice) (Table 2). There were no significant differences between automatic and reference T2 measurements in per-patient (51.9 msec \( \pm \) 5.5 vs 51.6 msec \( \pm \) 5.7; \( P = .06 \)), per-slice (51.7 msec \( \pm \) 7.0 vs 51.6 msec \( \pm \) 6.8; \( P = .08 \)), and per-segment (51.5 msec \( \pm \) 8.4 vs 51.6 msec \( \pm \) 8.4; \( P = .38 \)) analyses. There was a strong correlation between automated and reference T2 measurements in per-patient (\( R = 0.88, \) slope = 1.01, \( P < .0001 \)), per-slice (1310 slices: \( R = 0.83, \) slope = 1.00, \( P < .0001 \)), and per-segment (6899 segments: \( R = 0.75, \) slope = 0.99, \( P < .0001 \)) analyses (Fig 4, top row). The automatic and reference T2 values were in good agreement in per-patient \((0.3 \text{msec} \pm 2.9, 95\% \text{confidence interval [CI]}: -5.5 \text{msec}, 6.0 \text{msec})\), per-slice \((0.1 \text{msec} \pm 4.6, 95\% \text{CI}: -8.9 \text{msec}, 9.1 \text{msec})\), and per-segment \((0.0 \text{msec} \pm 6.5, 95\% \text{CI}: -12.8 \text{msec}, 12.7 \text{msec})\) analyses (Fig 4, bottom row). Exp#4 had a success rate of 237 of 284 (83%) (per-patient) and 1111 of 1420 (78%) (per-slice) for reconstruction of ECV maps (Table 2). Automatic and reference ECV measurements were comparable in per-patient (28.7% \( \pm \) 4.0 vs 29.0% \( \pm \) 4.0; \( P = .06 \)) and per-slice (28.7% \( \pm \) 4.3 vs 29.0% \( \pm \) 4.2; \( P = .05 \)) analyses but not in per-segment analysis (28.7% \( \pm \) 5.0 vs 29.0% \( \pm \) 5.0; \( P < .0001 \)). The automatic ECV measurements showed a strong correlation with the reference ECV in per-patient \((R = 0.91, \) slope = 0.99, \( P < .0001 \)), per-slice \((1101 \text{slices}: R = 0.84, \) slope = 0.99, \( P < .0001 \)), and per-segment \((6275 \text{segments}: R = 0.76, \) slope = 0.98, \( P < .0001 \)) analyses (Fig 5, top row). Automatic and reference ECV values were in good agreement in per-patient \((-0.3\% \pm 1.7, 95\% \text{CI}: -3.5\%, 3.0\%)\), per-slice \((-0.3\% \pm 2.5, 95\% \text{CI}: -5.1\%, 4.5\%)\), and per-segment ECV \((-0.4\% \pm 3.5, 95\% \text{CI}: -7.2\%, 6.5\%)\) analyses (Fig 5, bottom row).}

Table 2 summarizes the performance of various CNNs at automating T2 and ECV measurements. For T2 mapping, there was a strong correlation between the automated and reference T2 measurements among all three networks in per-patient \((R > 0.88, P < .0001)\), per-slice \((R > 0.83, P < .0001)\), and per-segment \((R > 0.75, P < .0001)\) analyses. Addition of T2-weighted images to the network trained by native T1-weighted images did not significantly improve success rate \((277 \text{of 284 [98\%] vs 276 \text{of 284 [97\%]}})\) or segmentation accuracy \((\text{Dice similarity coefficient: 0.87 \pm 0.07 vs 0.87 \pm 0.11})\). However, CNN\(_{\text{T1}}\)\(_{\text{T2}}\) had a lower success rate than the other two pretrained networks CNN\(_{\text{T1}}\)\(_{\text{nd}}\) and CNN\(_{\text{T2}}\)\(_{\text{nd}}\) presumably due to its smaller total number of training images \((945 \text{images for CNN\(_{\text{T1}}\)\(_{\text{nd}}\) vs 3465 \text{images for CNN\(_{\text{T2}}\)\(_{\text{nd}}\))}\). These results suggest that a network pretrained with native T1 mapping data can be used to segment T2-weighted images without compromising its performance. For ECV mapping, addition of postcontrast T1-weighted images to the network trained with only native T1-weighted images resulted in 4% improvement in success rate \((246 \text{of 284 [87\%] vs 237 of 284 [83\%]})\) and 0.03 improvement in segmentation accuracy \((0.87 \pm 0.08 \text{vs} 0.84 \pm 0.12)\). There was a strong correlation between automated and reference ECV measurements in all three CNNs in per-patient \((R > 0.91, P < .0001)\), per-slice \((R > 0.84, P < .0001)\), and per-segment \((R > 0.76, P < .0001)\) analyses (Table 2).

Table 3 summarizes the ability of CNN\(_{\text{T1}}\)\(_{\text{T2}}\) to identify patients with increased T2 (ie, T2 \( \geq \) 58 msec) or ECV (ie, ECV \( \geq \) 29%). The network’s sensitivity and specificity were 25 of 35 \((71.4\%)\) and 230 of 241 \((95.4\%)\) for T2 mapping, and 81 of 96 \((84.4\%)\) and \((127 \text{of 141 [90.1\%]})\) for ECV mapping. Errors in patient

Table 1: Performance Summary of Five Convolutional Neural Networks to Automate T2 and ECV Measurements

<table>
<thead>
<tr>
<th>Network</th>
<th>Initialization</th>
<th>Training Images</th>
<th>Data Size*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN(<em>{\text{nd}})(</em>{\text{T1}})</td>
<td>Random weights</td>
<td>Native T1 mapping</td>
<td>3465</td>
</tr>
<tr>
<td>CNN(<em>{\text{T1}})(</em>{\text{nd}})</td>
<td>Weights of CNN(<em>{\text{nd}})(</em>{\text{T1}})</td>
<td>T2 mapping</td>
<td>945</td>
</tr>
<tr>
<td>CNN(<em>{\text{T1}})(</em>{\text{nd}})</td>
<td>Weights of CNN(<em>{\text{nd}})(</em>{\text{T1}})</td>
<td>Postcontrast T1 mapping</td>
<td>1155</td>
</tr>
<tr>
<td>CNN(<em>{\text{nd}})(</em>{\text{T2}})</td>
<td>Random weights</td>
<td>T2 mapping</td>
<td>945</td>
</tr>
<tr>
<td>CNN(<em>{\text{nd}})(</em>{\text{T1}})</td>
<td>Random weights</td>
<td>Postcontrast T1 mapping</td>
<td>1155</td>
</tr>
</tbody>
</table>

Note.—Five different convolutional neural networks (CNNs) were investigated to evaluate their performance at automating T2 and extracellular volume (ECV) measurements. CNNs were initialized using different weighted parameters and trained using different image weightings and data sizes.

* Number of images in training.
identification occurred in patient images with parameter values close to the cutoff threshold; where the average values in the misidentified patients were less than 5 msec and less than 3% for T2 and ECV mapping, respectively.

Although the network was capable of delineating the myocardium with high accuracy in the majority of cases, failure was more likely in images with low signal-to-noise ratio or low myocardium-blood contrast (eight in T2 mapping and 47 in ECV mapping) (Fig 6). Although CNN segmentation achieved good performance in per-slice and per-patient analyses, per-segment analysis showed relatively larger errors in estimating T2 and ECV values. The majority of the segments with large measurement errors was located in the apical slices, where the myocardium boundaries are often difficult to identify or had severe image artifacts.

Discussion

Our study demonstrated the performance of transfer learning to automate T2 and ECV mapping analysis via a pretrained deep learning–based myocardial native T1 mapping analysis platform. The employed network yielded good performance, and the resulting T2 and ECV measurements agreed well with those from manual analyses. The employed transfer learning strategy proved effective, and networks with and without retraining showed comparable performance. Compared with transfer learning–based networks, networks trained from scratch achieved similar success rates for ECV mapping but showed lower success rates for T2 mapping. This may be due to the limited number of T2 mapping images in our training dataset, highlighting the value of transfer learning in the absence of large manually annotated datasets. In our study, from-scratch network training was performed using a manually segmented dataset of approximately 1000 images and employed two approaches to mitigate model overfitting: (a) a dropout layer (31) at the output of each network functional block (Figure E1 [supplement]) that randomly passed or blocked the processed data onto the next functional block, and (b) image augmentation to synthesize new training images by randomly translating, mirroring, and deforming training images. Although transfer learning produced superior network models, we expect the performance of from-scratch trained models to increase with the size of training set. However, we also note that, even with the availability of a larger manually segmented dataset, transfer learning can still be employed to allow efficient training of deeper models with a larger number of parameters.

It is also worth noting that we employed transfer learning mainly to mitigate the limitation...
Table 2: Evaluation Metrics for Automatic T2 and ECV Mapping by Network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Exp#1</th>
<th>Exp#2</th>
<th>Exp#3</th>
<th>Exp#4</th>
<th>Exp#5</th>
<th>Exp#6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>CNN\textsuperscript{nd} T1</td>
<td>CNN\textsuperscript{T1} T2</td>
<td>CNN\textsuperscript{nd} T2</td>
<td>CNN\textsuperscript{nd} T1</td>
<td>CNN\textsuperscript{T1} postT1</td>
<td>CNN\textsuperscript{nd} postT1</td>
</tr>
<tr>
<td>Map</td>
<td>T2</td>
<td>T2</td>
<td>T2</td>
<td>ECV</td>
<td>ECV</td>
<td>ECV</td>
</tr>
<tr>
<td>Refinement rate</td>
<td>328/12780 (2.6)</td>
<td>335/12780 (2.6)</td>
<td>518/12780 (4.1)</td>
<td>1067/31240 (3.4)</td>
<td>900/31240 (2.9)</td>
<td>899/31240 (2.9)</td>
</tr>
<tr>
<td>DSC*</td>
<td>0.87 ± 0.11</td>
<td>0.87 ± 0.07</td>
<td>0.88 ± 0.07</td>
<td>0.84 ± 0.12</td>
<td>0.87 ± 0.08</td>
<td>0.89 ± 0.07</td>
</tr>
<tr>
<td>Per patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>276/284 (97)</td>
<td>277/284 (98)</td>
<td>251/284 (88)</td>
<td>237/284 (83)</td>
<td>246/284 (87)</td>
<td>243/284 (86)</td>
</tr>
<tr>
<td>Slope</td>
<td>1.01 [1.00, 1.01]</td>
<td>1.07 [1.06, 1.08]</td>
<td>1.09 [1.08, 1.10]</td>
<td>0.98 [0.97, 1.00]</td>
<td>0.98 [0.97, 0.99]</td>
<td>0.99 [0.98, 1.00]</td>
</tr>
<tr>
<td>R</td>
<td>0.88 [0.85, 0.91]</td>
<td>0.84 [0.81, 0.88]</td>
<td>0.84 [0.80, 0.88]</td>
<td>0.89 [0.89, 0.93]</td>
<td>0.90 [0.90, 0.94]</td>
<td>0.91 [0.91, 0.94]</td>
</tr>
<tr>
<td>Per slice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>1329/1420 (94)</td>
<td>1312/1420 (92)</td>
<td>1236/1420 (87)</td>
<td>1111/1420 (78)</td>
<td>1151/1420 (81)</td>
<td>1141/1420 (80)</td>
</tr>
<tr>
<td>Slope</td>
<td>1.00 [0.99, 1.00]</td>
<td>1.07 [1.06, 1.07]</td>
<td>1.09 [1.08, 1.10]</td>
<td>0.98 [0.97, 0.99]</td>
<td>0.98 [0.97, 0.98]</td>
<td>0.99 [0.98, 0.99]</td>
</tr>
<tr>
<td>R</td>
<td>0.83 [0.81, 0.85]</td>
<td>0.83 [0.81, 0.84]</td>
<td>0.85 [0.83, 0.87]</td>
<td>0.84 [0.82, 0.86]</td>
<td>0.87 [0.85, 0.89]</td>
<td>0.87 [0.85, 0.89]</td>
</tr>
<tr>
<td>Per segment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.99 [0.99, 1.00]</td>
<td>1.07 [1.06, 1.07]</td>
<td>1.09 [1.08, 1.09]</td>
<td>0.98 [0.97, 0.98]</td>
<td>0.99 [0.98, 0.98]</td>
<td>0.99 [0.98, 0.99]</td>
</tr>
<tr>
<td>R</td>
<td>0.75 [0.74, 0.77]</td>
<td>0.77 [0.76, 0.79]</td>
<td>0.77 [0.75, 0.79]</td>
<td>0.76 [0.75, 0.77]</td>
<td>0.79 [0.78, 0.80]</td>
<td>0.80 [0.79, 0.81]</td>
</tr>
</tbody>
</table>

Note.—Table 2 includes the refinement rate of images using CNN segmentation (the 31 240 images for extracellular volume [ECV] included both native and postcontrast T1 mapping images), Dice similarity coefficient (DSC) of the automated segmented images, the success rate (SR) of map production, and the slope (S) and correlation coefficient (R) of the linear relationship between the automated and reference measurements. Data in parentheses for refinement rate, per-patient SR, and per-slice SR are percentages. Data in brackets are 95% confidence intervals.

* Data are means ± standard deviations for DSC.

Figure 4: Scatterplots and Bland-Altman plots of automated (using CNN\textsuperscript{nd} T1) and reference T2 values in per-patient, per-slice, and per-segment analyses. Automated T2 measurements showed a strong correlation with the reference T2 values in per-patient, per-slice, and per-segment analyses. Dashed lines in scatterplots represent the unity slope line. Automatic and reference T2 values were in good agreement in per-patient (0.3 msec ± 2.9 msec; 95% confidence interval [CI]: −5.5 msec to 6.0 msec), per-slice (0.1 msec ± 4.6 msec; 95% CI: −8.9 msec to 9.1 msec), and per-segment (0.0 msec ± 6.5 msec; 95% CI: −12.8 msec to 12.7 msec) analyses. Solid and dashed lines in Bland-Altman plots represent the bias and 6 1.96 standard deviation limits, respectively. CNN = convolutional neural network.
of small annotated datasets of T2-weighted and postcontrast T1-weighted images. However, direct application of CNNs to T2 and ECV mapping has saved computational resources required for training of a new network. Further saving in computational resources might be needed on some platforms, which might be achieved through other approaches such as attention transfer and may require further investigation.

In our study, CNN was only used to segment the myocardium on T1- and T2-weighted images acquired for parametric mapping. In our pipeline, the CNN was predominantly used to segment the left ventricle, and aligned and/or segmented regions of interest were conventionally fitted to calculate T1 or T2 values. The CNN could alternatively be applied over the entire pipeline or used only to segment the resulting T1 or T2 maps; however, these alternatives were not investigated in our study. We chose to decouple the fitting and segmentation aspects of the myocardial T1 and T2 mapping processing pipeline to rigorously investigate the performance of segmentation in the pipeline. While we use a network trained to segment T1-weighted images, other networks trained for other segmentation tasks, such as cine or perfusion, could potentially be used and should be further investigated. Moreover, having motion correction and parametric fitting steps separated in the proposed analysis pipeline can be advantageous.

Table 3: Performance Evaluation of Automated Identification of Patients with Increased T2 or ECV Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>T2</th>
<th>ECV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Patients</td>
<td>Mean T2 (msec)</td>
</tr>
<tr>
<td>True Negative: reference &lt; TH automated &lt; TH</td>
<td>230</td>
<td>50.2 ± 3.3 (49.5)</td>
</tr>
<tr>
<td>True Positive: reference ≥ TH automated ≥ TH</td>
<td>25</td>
<td>63.9 ± 5.5 (61.8)</td>
</tr>
<tr>
<td>False Positive: reference &lt; TH automated ≥ TH</td>
<td>11</td>
<td>59.0 ± 1.0 (58.6)</td>
</tr>
<tr>
<td>False Negative: reference ≥ TH automated &lt; TH</td>
<td>10</td>
<td>53.5 ± 4.5 (55.6)</td>
</tr>
</tbody>
</table>

Note.—Threshold (TH) of 58 msec for T2 mapping and 29% for extracellular volume (ECV) mapping was used. Data are means ± standard deviations, and data in parentheses are medians.
compared with previous methods (10,32) that combine both steps. For example, a separate step for parametric fitting allows flexibility in selecting the model type (e.g., two-parameter or three-parameter) without a need to re-perform the time-consuming motion correction step. Additionally, it allows exclusion of unsuccessfully segmented images from parametric fitting, thereby avoiding potential mapping errors.

Transfer learning–based segmentation of tissue mapping images is more challenging compared with segmentation in other sequences such as cine. Cine images acquired using balanced steady-state free precession sequences, for example, have high and temporally consistent myocardium-to-blood contrast ideal for efficient CNN training. Currently, most cardiac MRI analysis vendors offer automated cine segmentation for contouring endocardial and epicardial borders, and there are already several commercially available deep learning–based cine segmentation methods. On the other hand, to the best of our knowledge, there is no software available for automated analysis of T1, T2, or ECV mapping. Automated mapping analysis is challenged by the substantial variations in the myocardium-to-blood contrast caused by the different T1 or T2 weighting. At certain inversion or T2 preparation times, the myocardium-to-blood contrast is very low and usually impedes accurate myocardial segmentation. Also, there are currently numerous T1 and T2 mapping sequences that yield different contrasts. Different gadolinium-based contrast agents and acquisition times after contrast material administration also substantially impact the signal and contrast-to-noise ratio of postcontrast T1-weighted images. Learning image features from such diverse image contrast patterns requires large training datasets that might not be available, highlighting the benefit of transfer learning–based myocardium segmentation in tissue mapping images.

The current practice of quantifying T1, T2, or ECV varies across different centers. Septal measurements have been shown to have high reproducibility (33) and are often used in cardiac MRI tissue mapping (34,35). However, because diffuse fibrosis or inflammation could be regional (3,9), a septal measurement is insufficient. Myocardial tissue mapping sequences with whole heart coverage have been developed (7,9). However, manual analysis is time-consuming and operator dependent. The proposed automated analysis platform has the potential to facilitate adoption of whole-heart T1 or T2 mapping sequences for better characterization of tissue composition across different regions.

Our study had several limitations. All images were acquired using the same MRI scanner at 1.5 T. Because image quality varies with field strength and scanner type, further studies are warranted to evaluate the platform performance for myocardial tissue maps acquired at different field strengths and from different vendors. A preliminary experiment showed that the developed method may have the potential to automate T1 mapping using other sequences; e.g., MOLLI (modified Look-Locker inversion recovery) (Appendix E1 [supplement]). However, a rigorous study with a large dataset is needed to evaluate the performance of generalizing transfer learning to a wider spectrum of mapping sequences. Also, myocardial alignment in our study was performed by forward and backward transformation from Cartesian to polar coordinates, which could introduce interpolation errors to the computed maps. Although interpolation errors may be suppressed during analysis through averaging of the myocardium maps, further investigation is needed to determine their impact on the computed maps compared with other sources of error such as imaging artifacts and noise. We used manual analysis by an experienced observer as the reference standard. However, there is still intraobserver variability in the manual measurements (29). For per-segment analysis, the selection of insertion points was manually selected in the current analysis platform.

In conclusion, our CNN-based automated analysis platform, trained on T1 mapping data, yielded T2 and ECV values in good agreement with manual measurements without the need to retrain the network with corresponding T2 and ECV mapping data. This platform has the potential to fully automate myocardial tissue mapping, allowing standardization of data analysis.

**Acknowledgment:** The authors wish to thank Jennifer Rodriguez for the editorial corrections.

**Author contributions:** Guarantors of integrity of entire study, Y.Z., R.N.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all.
References


