



Article Fully Automated Regional Analysis of Myocardial T2* Values for Iron Quantification Using Deep Learning

Nicola Martini ^{1,2,3,†}, Antonella Meloni ^{1,2,†}, Vincenzo Positano ^{1,2}, Daniele Della Latta ^{2,3}, Petra Keilberg ¹, Laura Pistoia ¹, Anna Spasiano ⁴, Tommaso Casini ⁵, Angelica Barone ⁶, Antonella Massa ⁷, Andrea Ripoli ^{2,3} and Filippo Cademartiri ^{1,*}

- ¹ Department of Radiology, Fondazione G. Monasterio CNR-Regione Toscana, 56124 Pisa, Italy
- ² Unità Operativa Complessa di Bioingegneria, Fondazione G. Monasterio CNR-Regione Toscana, 56124 Pisa, Italy
- ³ Deep Health Unit, Fondazione G. Monasterio CNR-Regione Toscana, 56124 Pisa, Italy
- ⁴ Unità Operativa Semplice Dipartimentale Malattie Rare del Globulo Rosso, Azienda Ospedaliera di Rilievo Nazionale "A. Cardarelli", 80131 Napoli, Italy
- ⁵ Centro Talassemie ed Emoglobinopatie, Ospedale "Meyer", 50139 Firenze, Italy
- ⁶ Unità Operativa di Pediatria e Oncoematologia, Dipartimento Materno-Infantile,
- Azienda Ospedaliero-Universitaria di Parma, 43100 Parma, Italy
- ⁷ Servizio Trasfusionale, Ospedale "Giovanni Paolo II", 07026 Olbia, Italy
- * Correspondence: fcademartiri@ftgm.it; Tel.: +39-050-3152817
- † These authors contributed equally to this work.

Abstract: Cardiovascular magnetic resonance (CMR) T2* mapping is the gold standard technique for the assessment of iron overload in the heart. The quantitative analysis of T2* values requires the manual segmentation of T2* images, which is a time-consuming and operator-dependent procedure. This study describes a fully-automated method for the regional analysis of myocardial T2* distribution using a deep convolutional neural network (CNN). A CNN with U-Net architecture was trained to segment multi-echo T2*-weighted images in 16 sectors in accordance with the American Heart Association (AHA) model. We used images from 210 patients (three slices, 10 multi-echo images) with iron overload diseases to train and test the CNN. The performance of the proposed method was quantitatively evaluated on an independent holdout test set by comparing the segmentation accuracy of the CNN and the T2* values obtained by the automated method against ground-truth labels provided by two experts. Segmentation metrics and global and regional T2* values assessed by the proposed DL method closely matched those obtained by experts with excellent intraclass correlation in all myocardial sectors of the AHA model (ICC range [0.944, 0.996]). This method could be effectively adopted in the clinical setting for fast and accurate analysis of myocardial T2*.

Keywords: T2* mapping; cardiovascular magnetic resonance; segmentation; deep learning; myocardial iron overload; quantitative MRI

1. Introduction

Iron overload, designated as excess stores of iron in the body, is a surprisingly common condition affecting millions of people worldwide. It can result from inherited disorders of iron homeostasis characterized by increased intestinal iron absorption (primary hemochromatosis) and from chronic transfusions in patients with acquired and inherited anemias such as thalassemia and sickle cell disease (secondary hemochromatosis) [1]. Excess iron is deposited in the parenchyma of many tissues and, being toxic, can cause tissue damage and organ dysfunction, especially in the heart, liver, and pancreas [2]. In particular, iron-induced cardiomyopathy continues to be a leading cause of morbidity and mortality in patients with primary as well as secondary hemochromatosis [3]. If diagnosed and treated in its early stages, this condition is treatable and reversible [4]. Therefore, the quantification of myocardial iron overload (MIO) is the key to better patient management.



Citation: Martini, N.; Meloni, A.; Positano, V.; Latta, D.D.; Keilberg, P.; Pistoia, L.; Spasiano, A.; Casini, T.; Barone, A.; Massa, A.; et al. Fully Automated Regional Analysis of Myocardial T2* Values for Iron Quantification Using Deep Learning. *Electronics* **2022**, *11*, 2749. https:// doi.org/10.3390/electronics11172749

Academic Editor: Xiaojun Chen

Received: 27 July 2022 Accepted: 30 August 2022 Published: 1 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

To date, cardiovascular magnetic resonance (CMR) is the only technique able to provide a quantitative assessment of the cardiac iron burden in a non-invasive manner [5]. The presence of myocardial iron deposits causes microscopic magnetic field inhomogeneities and results in a reduction in all relaxation times (T1, T2, and T2*) [6], with the T2* technique representing the gold standard. The T2* value is calculated by fitting the CMR signal acquired at different echo times (TEs) with a proper exponential decay model [7]. Two validated approaches are generally used in the clinical setting: the single-slice and the multislice approach. In the single-slice approach, the T2* value is calculated in a single region of interest (ROI) delineated in the mid-ventricular septum [8], while the multislice approach has been developed for the regional analysis of the entire left ventricle (LV) [9]. The multislice approach allows the identification of an early and heterogeneous iron distribution that would remain otherwise undiscovered with a single measurement in the mid-ventricular septum [10], but it is more tedious, time-consuming, and operator-dependent. In fact, users are required to manually delineate the LV boundaries and identify the right ventricle (RV) insertion point to divide the LV into equiangular sectors, in accordance with the American Heart Association (AHA) model [11].

Deep learning (DL), a branch of artificial intelligence (AI), is well-suited to streamline this workflow, offering the promise of full automation and output consistency and repeatability. DL algorithms learn efficient features or patterns directly from the input data (raw data) and combine these features for classification without the need for prior knowledge and human intervention [12].

The most successful type of DL algorithms for image analysis are the convolutional neural networks (CNNs) [13]. The architecture of a typical CNN network is composed of multiple layers ("hidden layers"), such as convolution layers, pooling layers, and fully connected layers that, in turn, map the input image to the desired output while learning increasingly higher-level imaging features through a backpropagation algorithm [14]. Convolution layers are the core building blocks of the CNN that perform feature extraction by applying a set of filters to the image to produce spatially dependent features. Pooling layers perform downsampling operations to decrease the spatial size of the representation and the amount of computation and weights. In classification tasks, the extracted features are finally mapped by fully connected layers to the final outputs, such as the probabilities for each class [15]. Instead, in semantic segmentation tasks, fully convolutional networks (FCN) trained end-to-end, pixels-to-pixels, have shown to obtain the best results [16].

In the last few years, CNNs have made a tremendous impact on the entire workflow in CMR imaging [17], bringing unprecedented benefits in the areas of image acquisition [18], image reconstruction [19,20], image segmentation [21], and diagnostic evaluation [22]. It has been shown that CNNs can be used to process multiple CMR sequences (cine, late gadolinium enhancement, native and post-contrast T1), achieving robust segmentation results and leading to quantitative results (i.e., ejection fractions or T1 values) close to those obtained with manual segmentation [23–25].

To the best of our knowledge, no deep learning–based analysis platform has yet been developed for automating the global and regional analysis of myocardial T2* values. In this work, we sought to develop and validate an automatic method for the AHA model segmentation of T2* mapping using a deep convolutional neural network. The performance of the proposed method was quantitatively evaluated on an independent holdout test set by comparing the segmentation accuracy of the CNN and the T2* values obtained by the automated method against the ground-truth contours and measurements provided by expert operators.

2. Materials and Methods

2.1. Study Population

Images from 210 patients with iron overload diseases (106 females, 38.2 ± 12.8 years) were retrospectively studied. All of the patients were consecutively enrolled from 2015 to 2018 in the coordinator center of the Extension-Myocardial Iron Overload in Thalassemia

(E-MIOT) project (Pisa). E-MIOT is an Italian network composed of 66 thalassemia centers and 11 magnetic resonance imaging (MRI) sites that perform CMR exams using homogeneous, standardized, and validated procedures [26,27].

The study complied with the Declaration of Helsinki and was approved by the institutional ethics committee. All patients provided written informed consent to the protocol.

2.2. Image Acquisition and T2* Map Generation

CMR was performed on two clinical 1.5T scanners (Signa CVi or Signa Artist, GE Healthcare, Milwaukee, WI, USA) using a cardiac phased-array receiver surface coil for signal reception.

Three slices (basal, medial, and apical) of the LV in short-axis view were collected using an ECG-triggered T2* gradient–echo multi-echo sequence [9]. The images were acquired during the tele-diastolic phase to minimize cardiac motion [28]. The multi-echo sequence parameters were as follows: number of echo times, 10; first echo time, 2.0 ms; echo spacing, 2.26 ms; flip angle, 25° ; matrix, 192×256 pixels; field of view (FOV), 35×35 cm; bandwidth, 62.5 KHz; slice thickness, 8.0 mm; and views per segment, 6–8.

The fully automatic process for T2* map computation has already been described and validated [29]. Briefly, the T2* value for each pixel was calculated by fitting the corresponding MRI signal at the increasing TEs with a single exponential decay model:

$$S = S_0 e^{\frac{-1E}{T2*}}$$
(1)

where S indicates the signal intensity, S_0 is the signal intensity at TE = 0, and TE represents the echo times. The fitting was performed using the Levenberg–Marquardt algorithm. In patients with severe MIO, the signal decays quickly and becomes comparable to image noise, thus hampering the goodness and generating a high fitting error, computed as the normalized mean root square error between the measured MR signal and the fitted decay curve. If the fitting error was more than 5%, the algorithm discarded the last TE and performed the fitting again. The procedure was iterated until the error became <5% or the number of TEs became equal to three.

2.3. Data Preparation and Labeling

The acquired T2* images were analyzed by an expert CMR operator (P. K., >15 years of experience in T2* analysis) using a previously validated, custom-written IDL-based software (Hippo-MIOT[®]) to obtain ground-truth data [29]. Briefly, for each short-axis slice, the operator delineated the endocardial and epicardial borders of the LV in an image corresponding to the first or second echo time and defined a reference point in the anterior septal insertion of the right ventricle. The myocardium defined in the previous step was automatically segmented into equiangular segments starting from the reference point. According to the AHA model, the basal and medial slices were divided into 6 segments and the apical slice into 4 segments [11]. The T2* values in each segment was obtained by averaging the T2* values for all the pixels within the segment. The T2* values of each slice were obtained by averaging the T2* values of all myocardial pixels, regardless of the AHA sector. The ground-truth segmentation masks and T2* maps were stored in Hierarchical Data Format version 5 (HDF5) for DL model training and performance evaluation.

2.4. Deep Learning Segmentation

The segmentation of three slices into 16 AHA-based myocardial segments was translated into a single-slice multiclass semantic segmentation problem with the number of predicted classes equal to 7, i.e., the six myocardial segments and the background. To this end, the ground-truth masks of the apical slice were subdivided into 6 segments for the training, splitting the septal and lateral segments into two.

A fully convolutional neural network architecture (U-Net) [30] was adopted for segmentation. Our U-Net model was fed with full FOV single-slice raw T2* multi-echo images (array of $256 \times 256 \times 10$) and was trained to give as output the corresponding 7-classes

mask (Figure 1). Each block of our U-Net consisted of two convolutional layers with a kernel size of 5×5 (stride 1 and padding 1) with a 1×1 convolutional layer in between to reduce the depth of the feature maps and, in turn, reduce the total number of network parameters. The Rectified Linear Unit (ReLU) was used as an activation function, and batch normalization layers were inserted after each convolutional layer. Max pooling layers were used for downsampling, while transposed convolution layers (stride 2, padding 2) performed the upsampling of the feature maps to restore the original size of the input image. The output of the last convolutional layer was a $256 \times 256 \times 7$ array of scores representing the 7 segmented classes, which were converted to normalized probability values via a softmax function. Finally, the predicted segmentation mask was obtained by pixel-wise application of the *argmax* function on softmax probabilities.



Figure 1. The proposed U-Net model architecture for the automatic segmentation of multi-echo gradient-echo images.

2.5. Network Training

The proposed U-Net model was developed in Python using the TensorFlow and Keras framework. Network training was performed on a computer with the following specifications: 12-core Intel i7-7800 CPU 3.5 GHz with 64 GB RAM and an NVIDIA Titan Xp GPU with 12 GB of memory, and running the Linux operating system.

The whole dataset of the T2* multi-echo images was split into a training set (N = 510 slices), validation set (N = 60 slices), and test set (N = 60 slices) for performance evaluation. During training, random rotations (range $[-90^\circ, +90^\circ]$) and scaling transformations (range [0.8, 1.2]) were used for data augmentation. A linear interpolation was used for transforming input images, while nearest-neighbor interpolation was applied to the ground-truth label masks.

Adam optimizer with initial learning rate of 0.001 ($\beta 1 = 0.9$, $\beta 2 = 0.999$, epsilon = 1e-7) was used with a batch size of 16. The weighted cross-entropy (WCE) loss was chosen as a loss function to address the issue of class imbalance. The class weights were set inversely proportional to the frequencies of each class label in the training data. Iterative training and validation steps were carried out on 2000 epochs, with 32 steps for each epoch, performing each validation at the end of each epoch. The final model was selected as the one achieving the best performance on the validation set.

2.6. Evaluation of the Model Performance

The performance of the automated method was evaluated in two ways: using commonly used metrics for the assessment of the segmentation accuracy and using quantitative measures (segmental and global T2* values) derived from the segmentations.

2.6.1. Segmentation Accuracy Assessment

The Dice similarity coefficient, average perpendicular distance, and Hausdorff distance were used as metrics of segmentation performance.

The Dice similarity coefficient (DSC) evaluates the overlap between the automated or predicted segmentation mask A and the manual segmentation mask B, and it is defined as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{2}$$

DSC ranges between 0 and 1, with 0 denoting no overlap and 1 denoting perfect agreement. The higher the DSC metric, the better the agreement. Following previous studies, a DSC > 0.7 was defined as substantial agreement, between 0.4 and 0.7 as moderate agreement, and <0.4 as large variation [31,32].

The average perpendicular distance (APD) distance and Hausdorff distance (HD) evaluate, respectively, the mean and the maximum of the minimum surface distances between the segmentation contours ∂A and ∂B . They are defined as:

$$APD = \frac{1}{2|\partial A|} \sum_{p \in \partial A} d(p, \partial B) + \frac{1}{2|\partial B|} \sum_{q \in \partial B} d(q, \partial A)$$
(3)

$$HD = max \left(max_{p \in \partial A} d(p, \partial B), max_{q \in \partial B} d(q, \partial A) \right)$$
(4)

where $d(p, \partial)$ denotes the minimal distance from point p to contour ∂ . These distances indicate how much the contours should be quantitatively modified relative to the ground truth. The lower the distance metric, the better the agreement.

In addition to these commonly used segmentation metrics, we additionally proposed a custom metric called Angular Error (AE) for the assessment of the correct orientation of each myocardial segment. The AE was calculated as:

$$AE = \cos^{-1}\left(\frac{b \cdot a}{\|b\|\|a\|}\right) = \cos^{-1}\left(\frac{\overrightarrow{CB} \cdot \overrightarrow{CA}}{\|\overrightarrow{CB}\|\|\overrightarrow{CA}\|}\right)$$
(5)

where b and a denote the vectors between the centroids *B* and *A* of the predicted and the manual segmental masks, respectively, and the center of the myocardium (point *C*). The lower the AE metric, the better the agreement.

The DSC, APD, and HD metrics were calculated for each myocardial segment separately, for the entire myocardial wall in the three slices (in the following denoted by Myoc) and in the left ventricle cavities (denoted by LV).

2.6.2. Accuracy of T2* Quantification

All of the data were analyzed and visualized using Python 3.6 packages (NumPy, SciPy, Pandas, Matplotlib, Seaborn). The continuous variables are reported as mean \pm standard deviation (SD).

The correlation between the segmental and global T2* values obtained with the two different approaches (manual vs automatic segmentation) was evaluated with the Spearman test since the T2* values showed a non-normal distribution. The summary data were displayed using scatter plots with regression lines.

The coefficient of variation (CoV) was obtained as the ratio of the SD of the half mean square of the differences between the repeated values to the general mean. A CoV < 10% was considered good. The ICC was obtained from a two-way random effects model with measures of absolute agreement. An ICC \geq 0.75 was considered excellent, between 0.40 and 0.75 fair to good, and <0.40 poor [33].

The agreement between the DL algorithm and the manual operator was determined by the Bland–Altman technique, plotting the difference versus the average of the variables. Bias was the mean of the difference between the two methods, and agreement was the mean \pm 1.96 SD. In all of the tests, two-tailed *p* < 0.05 was considered statistically significant.

2.6.3. Assessment of Validity

The "conservative" value of 20 ms is commonly used as the lower limit of normal for the segmental and global heart T2* values at 1.5T [5].

The global T2* values obtained with manual segmentation were considered the standard of reference and were used to categorize patients into two clinically relevant groups: MIO (T2* < 20 ms) and no MIO (T2* \geq 20 ms). So, it was detected if classifications obtained for fully-automated global heart T2* values were true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The sensitivity, specificity, positive and negative predictive values (PPV and NPV), and accuracy were calculated.

2.6.4. Inter-Observer Agreement

The images of the test set were also analyzed by a second observer to evaluate the interobserver agreement and to compare the inter-observer variability with the performance of the automatic method. The images were presented in random order to another operator (V.P., >15 years of experience), blinded to the results obtained by the first observer. The segmentation metrics (DSC, APD, HD, and AE) and quantitative T2* values were evaluated between the pair of observers and compared against the DL method.

3. Results

Performance of U-Net

The proposed deep learning method was able to provide labels for all the 320 myocardial segments of the test data (20 patients \times 16 segments). Figure 2 illustrates the predicted segmentation of the LV according to the 16-segment AHA model. It shows that the automated segmentation agrees well with the manual segmentation of the clinical expert.



Figure 2. Representative examples of 16-segment AHA model myocardial segmentation from gradient-echo multi-echo T2* images in two testing cases: (**a**) subject with normal T2* values and (**b**) patient with iron overload. Ground-truth label masks (upper panels) and U-Net-based segmentations (lower panels) are shown for each of the three myocardial slices. Global T2* derived from the segmentation and Dice similarity coefficients (DSC) are also indicated.

The segmentation metrics evaluated in the test set (Table 1) demonstrated excellent agreement between the manual and fully-automated segmentations obtained by the U-Net in terms of volume overlap (mean Dice 0.771 [CI 0.755–0.787]), surface distance (APD = 0.61 mm [CI 0.53–0.69], HD = 4.30 mm [CI 3.95–4.65]), and angulation error (AE = 3.97° [CI 3.48–4.47]) across all segments.

Table 1. Dice similarity coefficients (DSC), average perpendicular distance (APD), Hausdorff distance (HD), and angulation error (AE) between automated and manual segmentations calculated for each of the 16 sectors of the AHA model, for the myocardial wall (Myoc) and for LV cavity (LV).

Segment/Myoc/LV	DSC	APD [mm]	HD [mm]	AE [°]
1: basal anterior	0.795 ± 0.086	0.42 ± 0.29	3.25 ± 1.60	3.44 ± 3.78
2: basal anteroseptal	0.794 ± 0.081	0.41 ± 0.26	3.68 ± 1.72	2.81 ± 3.22
3: basal inferoseptal	0.805 ± 0.084	0.37 ± 0.24	3.31 ± 1.86	2.98 ± 3.17
4: basal inferior	0.807 ± 0.101	0.38 ± 0.31	3.11 ± 1.34	2.36 ± 2.59
5: basal inferolateral	0.767 ± 0.191	0.57 ± 0.88	3.44 ± 2.28	3.90 ± 3.46
6: basal anterolateral	0.741 ± 0.186	0.63 ± 0.76	3.39 ± 1.93	3.27 ± 2.57
7: medium anterior	0.746 ± 0.157	0.57 ± 0.52	3.87 ± 1.86	3.87 ± 4.91
8: medium anteroseptal	0.772 ± 0.096	0.44 ± 0.28	3.39 ± 1.06	3.49 ± 3.19
9: medium inferoseptal	0.787 ± 0.145	0.44 ± 0.52	3.32 ± 1.45	3.31 ± 3.31
10: medium inferior	0.764 ± 0.146	0.53 ± 0.53	3.58 ± 1.77	3.45 ± 4.51
11: medium inferolateral	0.752 ± 0.160	0.61 ± 0.70	4.04 ± 2.32	3.27 ± 4.94
12: medium anterolateral	0.754 ± 0.203	0.64 ± 0.93	3.76 ± 2.38	4.23 ± 5.39
13: apical anterior	0.762 ± 0.129	0.54 ± 0.48	4.09 ± 2.05	5.84 ± 5.65
14: apical septal	0.775 ± 0.096	0.47 ± 0.35	4.06 ± 1.73	5.01 ± 5.78
15: apical inferior	0.762 ± 0.133	0.53 ± 0.50	4.11 ± 2.11	5.56 ± 5.48
16: apical lateral	0.730 ± 0.168	0.68 ± 0.68	4.43 ± 2.43	7.08 ± 6.12
Myoc: basal	0.819 ± 0.087	1.94 ± 0.92	13.18 ± 4.78	3.13 ± 2.27
Myoc: mid-ventricular	0.802 ± 0.095	2.11 ± 0.96	13.31 ± 5.81	3.60 ± 3.77
Myoc: apical	0.802 ± 0.090	2.12 ± 1.00	13.58 ± 6.51	5.87 ± 4.35
LV: basal	0.927 ± 0.032	2.44 ± 1.73	$\overline{13.45\pm8.39}$	n.a.
LV: mid-ventricular	$\overline{0.907\pm0.044}$	3.06 ± 2.01	16.32 ± 7.32	n.a.
LV: apical	0.911 ± 0.043	2.24 ± 1.29	13.23 ± 5.77	n.a

In the whole myocardial wall, the Dice values of the three slices were all above 0.8. The Dice and angulation error metrics showed better values for basal than mid-ventricular and apical slices, with the AE being significantly higher in the apical slice (basal vs apical: 3.13 ± 2.33 vs. 5.87 ± 4.46 , p < 0.0001; mid-ventricular vs. apical: 3.60 ± 3.87 vs. 5.87 ± 4.46 , p < 0.05).

In the LV cavity, the Dice values of the three slices were all above 0.9, and the APD values were all less than 3 mm.

In the test cohort, a strong correlation between the segmental and global T2* values obtained with the manual and the DL-based methods was obtained (Table 2). The CoV was always <10%, and the ICC was excellent.

Segment	Bland Altman Bias (Limits of Agreement) [ms]	Correlation (R; <i>p</i> -Value)	ICC (95%CI)	CoV (%)
1: basal anterior	-0.97 (-5.15 to 3.1)	R = 0.955; $p < 0.0001$	0.994 (0.983–0.998)	5.42
2: basal anteroseptal	-0.38 (-4.13 to 3.37)	R = 0.977; $p < 0.0001$	0.996 (0.990–0.998)	4.12
3: basal inferoseptal	0.09 (-3.64 to 3.83)	R = 0.982; $p < 0.0001$	0.996 (0.990–0.998)	3.97
4: basal inferior	0.10 (-2.63 to 2.83)	R = 0.992; $p < 0.0001$	0.997 (0.993–0.999)	3.42
5: basal inferolateral	1.11 (-5.63 to 7.85)	R = 0.919; p < 0.0001	0.971 (0.928–0.989)	9.99
6: basal anterolateral	-0.09 (-4.05 to 3.86)	R = 0.953; $p < 0.0001$	0.994 (0.985–0.998)	4.56
7: medium anterior	-1.09 (-7.19 to 5.01)	R = 0.962; p < 0.0001	0.987 (0.966–0.995)	8.37
8: medium anteroseptal	-0.64 (-4.45 to 3.17)	R = 0.992; $p < 0.0001$	0.995 (0.987–0.998)	4.67
9: medium inferoseptal	-0.97 (-4.28 to 2.35)	R = 0.995; $p < 0.0001$	0.996 (0.986–0.999)	4.17
10: medium inferior	-1.46 (-5.57 to 2.63)	R = 0.988; $p < 0.0001$	0.992 (0.964–0.997)	6.77
11: medium inferolateral	-0.74 (-6.02 to 4.55)	R = 0.961; $p < 0.0001$	0.986 (0.966–0.995)	7.64
12: medium anterolateral	-1.03 (-5.54 to 3.48)	R = 0.979; $p < 0.0001$	0.990 (0.974–0.996)	6.32
13: apical anterior	-0.08 (-5.37 to 5.21)	R = 0.982; $p < 0.0001$	0.992 (0.979–0.997)	6.71
14: apical septal	-0.41 (-3.68 to 2.86)	R = 0.991; $p < 0.0001$	0.997 (0.993–0.999)	3.87
15: apical inferior	-0.09 (-5.26 to 5.07)	R = 0.991; $p < 0.0001$	0.993 (0.982–0.997)	6.61
16: apical lateral	0.92 (-4.65 to 6.49)	R = 0.961; $p < 0.0001$	0.988 (0.969–0.995)	6.93
Global	-0.36 (-2.14 to 1.42)	R = 0.992; p < 0.0001	0.998 (0.996–0.999)	0.233

Table 2. Agreement between T2* values obtained with the manual and the DL-based methods in the test cohort.

There was a strong correlation between the automated and manual T2* measurements in the per-slice (Figure 3a) and per-segment (all together; Figure 3c) analyses. The Bland–Altman plot (Figure 3b,d) showed narrow limits of agreement (per-slice: 3 ms; per-sector: 5 ms) between the two assessments with no bias (<1 ms).



Figure 3. Cont.



Figure 3. (a,c) Scatter plots demonstrating a high correlation between fully-automated DL method and manually processed segmental T2* values. The segments of the three slices (basal, mid-ventricular, apical) are indicated by different markers. (b,d) Bland–Altman plot showing the excellent agreement between the manual and the proposed DL method for the regional quantitative analysis of T2*. Middle line denotes mean. Dashed lines denote \pm 1.96 standard deviations.

Table 3 shows the efficacy of the U-NET in differentiating the normal from pathological T2* values. The sensitivity was under 80% for three myocardial segments, while specificity was always very strong (high 90's). The accuracy was very high (high 90's). The global heart T2* value was always correctly classified.

	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	Accuracy (%)
1: basal anterior	100	100	100	100	100
2: basal anteroseptal	100	100	100	100	100
3: basal inferoseptal	100	100	100	100	100
4: basal inferior	80.00	100	100	93.75	95.00
5: basal inferolateral	85.71	100	100	92.86	95.00
6: basal anterolateral	100	100	100	100	100
7: medium anterior	75.00	100	100	85.71	90.00
8: medium anteroseptal	66.67	100	100	87.50	90.00
9: medium inferoseptal	100	100	100	100	100
10: medium inferior	71.43	100	100	86.67	90.00
11: medium inferolateral	85.71	92.31	85.71	92.31	90.00
12: medium anterolateral	83.33	100	100	93.33	95.00
13: apical anterior	85.71	100	100	92.86	95.00
14: apical septal	80.00	100	100	93.75	95.00
15: apical inferior	85.71	100	100	92.86	95.00
16: apical lateral	100	94.12	75.00	100	95.00
global	100	100	100	100	100

Table 3. Performance evaluation of automated identification of patients with reduced T2*.

The analysis of the inter-observer variability highlighted that the DL-based segmentation compares well with both operators. The DSC, APD, and HD segmentation metrics showed a better agreement between the DL-based method and each individual observer than the agreement between the two observers, as shown in Figure 4a. As a result, the regional distribution of the T2* values exhibited small differences (<3 ms) associated with excellent intra-class correlation coefficients (ICC) in all the myocardial segments (interobserver ICC), as displayed in Figure 4b.



Figure 4. Inter-observer variability of the two experts compared to the fully-automated method. (a) Boxplot of the segmentation metrics. (b) Bull's-eye plots showing the regional distribution of segmental T2* differences according to the AHA model (left panels) and ICC between observers and the DL-based method (right panels).

4. Discussion

This paper presents a deep neural network-based workflow for automated myocardial segmentation and reporting of the AHA 16 sector model for pixel-wise T2* mapping. To the best of our knowledge, this is the first study on the application of deep learning methods for the global and regional quantification of T2* in the heart.

The CNN used in this study was based on the U-Net architecture and trained using a weighted loss function. A variety of network architectures and losses can be used for segmentation in medical imaging, and the best solution may vary for different applications. A comprehensive overview of these methods can be found in [34].

The proposed CNN was able to provide fully-automated segmentation of the multiecho gradient-echo images with segmentation metrics in line with previous studies applied to other CMR acquisitions [35–37]. It is worth noting that, differently from cardiac cine analysis, for T2* quantification the endocardial and epicardial contours are usually traced in the myocardial midwall, i.e., with proper margins from the blood cavities, to avoid signal contamination from the blood. This might explain the lower Dice values (0.80–0.82) in the myocardial compared to some cardiac cine studies (0.87–0.94) [25,35], while our Dice coefficients in the LV cavity (0.90–0.93) well matched previous state-of-the-art techniques (0.90–0.94) [25,38]. Most interestingly, our inter-observer analysis showed a better agreement between the CNN segmentations and those provided by each individual observer compared to the agreement between the two experts, despite the fact that the second observer did not provide any training labels.

The key finding of our study is that global and regional T2* values assessed by the proposed DL method closely matched those obtained by expert CMR readers. The

Bland–Altman analysis demonstrated a 95% confidence interval for the global T2* of 3 ms compared to the manual measurements, which is more than acceptable for a fullyautomated tool that does not require any user intervention. The agreement is worse for high T2* values since, due to technical constraints (maximum gradient-echo time), the T2* quantification loses accuracy and precision for cardiac T2* values longer than 20 ms [39]. Importantly, the developed DL method allowed us to achieve high diagnostic reliability in the identification of MIO (100% for global heart T2* values). The sensitivity was under 80% only for three myocardial segments, probably due to partial contamination of the blood pool in the myocardial segmentation of the DL method.

The inter-observer analysis revealed an excellent intraclass correlation between the DL method and each of the observers in all myocardial sectors of the AHA model (ICC range [0.944, 0.996]). This finding is a further demonstration that deep learning tools can reach the same accuracy as human experts with the advantage of eliminating the inter-observer variability since these models, once trained, produce purely deterministic outputs. This advantage can lead to improved precision in test-retest scans and, ultimately, can increase diagnostic and prognostic performance [40].

In terms of speed, our CNN can process three short-axis slices in less than one second, thus enabling a fully-automated T2* analysis that can be readily incorporated into the scanners immediately after image acquisition.

Nevertheless, our study has several limitations. First, all of the CMR images of this study were acquired with two MR scanners from a single vendor and with the same field strength (1.5T). Second, all T2* images are collected with the same bright-blood T2* mapping sequence. However, other sequences, such as the black-blood T2* technique, have been proposed [41]. Given the different contrast between the blood cavity and the myocardium, a re-training of the network or a transfer learning procedure [38] should be performed to adapt the present CNN to these types of images. Third, our CNN took as input all the ten multi-echo images of the gradient-echo acquisition. Alternative approaches could use as input a single image or even the T2* map directly.

5. Conclusions

We proposed a deep learning approach for the automatic assessment of global and regional T2*. The segmentation results showed a close agreement with manually annotated masks. No significant differences in the segmental T2* values were found compared to the manual measurements. This method could be effectively implemented in the clinical arena not only for faster, accurate, and quality-controlled analysis in trained centers but also for a fast implementation of new T2* CMR sites, in particular in regions of the world with economic constraints.

Author Contributions: N.M.: conceptualization, network training, data analysis, manuscript writing. A.M. (Antonella Meloni): conceptualization, formal analysis, writing—original draft preparation. V.P. and D.D.L.: conceptualization, methodology, software. P.K.: image analysis. L.P.: data curation. A.S., T.C., A.B. and A.M. (Antonella Massa): data collection. A.R.: methodology. F.C.: supervision. All authors assisted with interpretation, commented on drafts of the manuscript, and approved the final version. All authors have read and agreed to the published version of the manuscript.

Funding: The E-MIOT project receives "no-profit support" from industrial sponsorships (Chiesi Farmaceutici S.p.A. and Bayer). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Area Vasta Nord Ovest (protocol code 56664, date of approval 8 October 2015).

Informed Consent Statement: Informed consent was obtained from all patients involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: We would like to thank all the colleagues involved in the E-MIOT project (https://emiot.ftgm.it/) and all patients for their cooperation.

Conflicts of Interest: D.D.L. is presently an employee of Terarecon Inc.; his collaboration to the present study occurred before its present affiliation, his contribution to this article reflects entirely and only his own expertise on the matter, and he declares no competing financial interests related to the present article. All the other authors do not have competing interests to disclose concerning the present manuscript.

References

- Al-Tubaikh, J.A. (Ed.) Hemosiderosis and Hemochromatosis. In *Internal Medicine*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 290–292, ISBN 978-3-642-03708-5.
- Pilo, F.; Cilloni, D.; Della Porta, M.G.; Forni, G.L.; Piperno, A.; Santini, V.; Angelucci, E. Iron-Mediated Tissue Damage in Acquired Ineffective Erythropoiesis Disease: It's More a Matter of Burden or More of Exposure to Toxic Iron Form? *Leuk. Res.* 2022, 114, 106792. [CrossRef] [PubMed]
- 3. Udani, K.; Chris-Olaiya, A.; Ohadugha, C.; Malik, A.; Sansbury, J.; Paari, D. Cardiovascular Manifestations in Hospitalized Patients with Hemochromatosis in the United States. *Int. J. Cardiol.* **2021**, *342*, 117–124. [CrossRef] [PubMed]
- Kremastinos, D.T.; Farmakis, D. Iron Overload Cardiomyopathy in Clinical Practice. *Circulation* 2011, 124, 2253–2263. [CrossRef] [PubMed]
- Pennell, D.J.; Udelson, J.E.; Arai, A.E.; Bozkurt, B.; Cohen, A.R.; Galanello, R.; Hoffman, T.M.; Kiernan, M.S.; Lerakis, S.; Piga, A.; et al. Cardiovascular Function and Treatment in β-Thalassemia Major: A Consensus Statement from the American Heart Association. *Circulation* 2013, *128*, 281–308. [CrossRef] [PubMed]
- Wood, J.C.; Otto-Duessel, M.; Aguilar, M.; Nick, H.; Nelson, M.D.; Coates, T.D.; Pollack, H.; Moats, R. Cardiac Iron Determines Cardiac T2*, T2, and T1 in the Gerbil Model of Iron Cardiomyopathy. *Circulation* 2005, 112, 535–543. [CrossRef] [PubMed]
- Meloni, A.; Rienhoff, H.Y.; Jones, A.; Pepe, A.; Lombardi, M.; Wood, J.C. Cardiac R2* Values Are Independent of the Image Analysis Approach Employed: Algorithms for Cardiac R2* Assessment. *Magn. Reson. Med.* 2014, 72, 485–491. [CrossRef] [PubMed]
- Westwood, M.; Anderson, L.J.; Firmin, D.N.; Gatehouse, P.D.; Charrier, C.C.; Wonke, B.; Pennell, D.J. A Single Breath-Hold Multiecho T2* Cardiovascular Magnetic Resonance Technique for Diagnosis of Myocardial Iron Overload. *J. Magn. Reson. Imaging* 2003, 18, 33–39. [CrossRef]
- Pepe, A.; Positano, V.; Santarelli, M.F.; Sorrentino, F.; Cracolici, E.; De Marchi, D.; Maggio, A.; Midiri, M.; Landini, L.; Lombardi, M. Multislice Multiecho T2* Cardiovascular Magnetic Resonance for Detection of the Heterogeneous Distribution of Myocardial Iron Overload. J. Magn. Reson. Imaging 2006, 23, 662–668. [CrossRef]
- Meloni, A.; Positano, V.; Pepe, A.; Rossi, G.; Dell'Amico, M.; Salvatori, C.; Keilberg, P.; Filosa, A.; Sallustio, G.; Midiri, M.; et al. Preferential Patterns of Myocardial Iron Overload by Multislice Multiecho T*2 CMR in Thalassemia Major Patients. *Magn. Reson. Med.* 2010, 64, 211–219. [CrossRef]
- Cerqueira, M.D.; Weissman, N.J.; Dilsizian, V.; Jacobs, A.K.; Kaul, S.; Laskey, W.K.; Pennell, D.J.; Rumberger, J.A.; Ryan, T.; Verani, M.S.; et al. Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart. A Statement for Healthcare Professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* 2002, 105, 539–542. [CrossRef]
- 12. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. Neural Netw. 2015, 61, 85–117. [CrossRef] [PubMed]
- Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* 2017, 42, 60–88. [CrossRef] [PubMed]
- Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial Intelligence in Radiology. *Nat. Rev. Cancer* 2018, 18, 500–510. [CrossRef] [PubMed]
- 15. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]
- 16. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
- Leiner, T.; Rueckert, D.; Suinesiaputra, A.; Baeßler, B.; Nezafat, R.; Išgum, I.; Young, A.A. Machine Learning in Cardiovascular Magnetic Resonance: Basic Concepts and Applications. J. Cardiovasc. Magn. Reson. Off. J. Soc. Cardiovasc. Magn. Reson. 2019, 21, 61. [CrossRef]
- 18. Blansit, K.; Retson, T.; Masutani, E.; Bahrami, N.; Hsiao, A. Deep Learning-Based Prescription of Cardiac MRI Planes. *Radiol. Artif. Intell.* **2019**, *1*, e180069. [CrossRef]
- Qi, H.; Hajhosseiny, R.; Cruz, G.; Kuestner, T.; Kunze, K.; Neji, R.; Botnar, R.; Prieto, C. End-to-End Deep Learning Nonrigid Motion-Corrected Reconstruction for Highly Accelerated Free-Breathing Coronary MRA. *Magn. Reson. Med.* 2021, *86*, 1983–1996. [CrossRef]
- Sandino, C.M.; Lai, P.; Vasanawala, S.S.; Cheng, J.Y. Accelerating Cardiac Cine MRI Using a Deep Learning-Based ESPIRiT Reconstruction. *Magn. Reson. Med.* 2021, 85, 152–167. [CrossRef]

- Bernard, O.; Lalande, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.-A.; Cetin, I.; Lekadir, K.; Camara, O.; Gonzalez Ballester, M.A.; et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans. Med. Imaging* 2018, 37, 2514–2525. [CrossRef]
- Martini, N.; Aimo, A.; Barison, A.; Della Latta, D.; Vergaro, G.; Aquaro, G.D.; Ripoli, A.; Emdin, M.; Chiappino, D. Deep Learning to Diagnose Cardiac Amyloidosis from Cardiovascular Magnetic Resonance. *J. Cardiovasc. Magn. Reson. Off. J. Soc. Cardiovasc. Magn. Reson.* 2020, 22, 84. [CrossRef] [PubMed]
- Romaguera, L.V.; Romero, F.P.; Fernandes Costa Filho, C.F.; Fernandes Costa, M.G. Myocardial Segmentation in Cardiac Magnetic Resonance Images Using Fully Convolutional Neural Networks. *Biomed. Signal Process. Control* 2018, 44, 48–57. [CrossRef]
- Fadil, H.; Totman, J.J.; Hausenloy, D.J.; Ho, H.-H.; Joseph, P.; Low, A.F.-H.; Richards, A.M.; Chan, M.Y.; Marchesseau, S. A Deep Learning Pipeline for Automatic Analysis of Multi-Scan Cardiovascular Magnetic Resonance. *J. Cardiovasc. Magn. Reson.* 2021, 23, 47. [CrossRef] [PubMed]
- Bai, W.; Sinclair, M.; Tarroni, G.; Oktay, O.; Rajchl, M.; Vaillant, G.; Lee, A.M.; Aung, N.; Lukaschuk, E.; Sanghvi, M.M.; et al. Automated Cardiovascular Magnetic Resonance Image Analysis with Fully Convolutional Networks. J. Cardiovasc. Magn. Reson. Off. J. Soc. Cardiovasc. Magn. Reson. 2018, 20, 65. [CrossRef]
- Ramazzotti, A.; Pepe, A.; Positano, V.; Rossi, G.; De Marchi, D.; Brizi, M.G.; Luciani, A.; Midiri, M.; Sallustio, G.; Valeri, G.; et al. Multicenter Validation of the Magnetic Resonance T2* Technique for Segmental and Global Quantification of Myocardial Iron. J. Magn. Reson. Imaging 2009, 30, 62–68. [CrossRef] [PubMed]
- Meloni, A.; Ramazzotti, A.; Positano, V.; Salvatori, C.; Mangione, M.; Marcheschi, P.; Favilli, B.; De Marchi, D.; Prato, S.; Pepe, A.; et al. Evaluation of a Web-Based Network for Reproducible T2* MRI Assessment of Iron Overload in Thalassemia. *Int. J. Med. Inf.* 2009, 78, 503–512. [CrossRef]
- Positano, V.; Pepe, A.; Santarelli, M.F.; Scattini, B.; De Marchi, D.; Ramazzotti, A.; Forni, G.; Borgna-Pignatti, C.; Lai, M.E.; Midiri, M.; et al. Standardized T2* Map of Normal Human Heartin Vivo to Correct T2* Segmental Artefacts. NMR Biomed. 2007, 20, 578–590. [CrossRef]
- Positano, V.; Meloni, A.; Santarelli, M.F.; Gerardi, C.; Bitti, P.P.; Cirotto, C.; De Marchi, D.; Salvatori, C.; Landini, L.; Pepe, A. Fast Generation of T2* Maps in the Entire Range of Clinical Interest: Application to Thalassemia Major Patients. *Comput. Biol. Med.* 2015, 56, 200–210. [CrossRef]
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015. In Proceedings of the MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015.
- 31. Zijdenbos, A.P.; Dawant, B.M.; Margolin, R.A.; Palmer, A.C. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. *IEEE Trans. Med. Imaging* **1994**, *13*, 716–724. [CrossRef]
- Carillo, V.; Cozzarini, C.; Perna, L.; Calandra, M.; Gianolini, S.; Rancati, T.; Spinelli, A.E.; Vavassori, V.; Villa, S.; Valdagni, R.; et al. Contouring Variability of the Penile Bulb on CT Images: Quantitative Assessment Using a Generalized Concordance Index. *Int. J. Radiat. Oncol. Biol. Phys.* 2012, *84*, 841–846. [CrossRef]
- 33. Rosner, B. Multisample Inference. In Fundamentals of Biostatistics; Duxbury: Pacific Grove, CA, USA, 2000; pp. 511–576.
- 34. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep Learning for Cardiac Image Segmentation: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [CrossRef] [PubMed]
- 35. Avendi, M.R.; Kheradvar, A.; Jafarkhani, H. A Combined Deep-Learning and Deformable-Model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI. *Med. Image Anal.* **2016**, *30*, 108–119. [CrossRef] [PubMed]
- Bratt, A.; Kim, J.; Pollie, M.; Beecy, A.N.; Tehrani, N.H.; Codella, N.; Perez-Johnston, R.; Palumbo, M.C.; Alakbarli, J.; Colizza, W.; et al. Machine Learning Derived Segmentation of Phase Velocity Encoded Cardiovascular Magnetic Resonance for Fully Automated Aortic Flow Quantification. J. Cardiovasc. Magn. Reson. Off. J. Soc. Cardiovasc. Magn. Reson. 2019, 21, 1. [CrossRef]
- 37. Xue, H.; Davies, R.H.; Brown, L.A.E.; Knott, K.D.; Kotecha, T.; Fontana, M.; Plein, S.; Moon, J.C.; Kellman, P. Automated Inline Analysis of Myocardial Perfusion MRI with Deep Learning. *Radiol. Artif. Intell.* **2020**, *2*, e200009. [CrossRef]
- Ankenbrand, M.J.; Lohr, D.; Schlötelburg, W.; Reiter, T.; Wech, T.; Schreiber, L.M. Deep Learning-based Cardiac Cine Segmentation: Transfer Learning Application to 7T Ultrahigh-field MRI. *Magn. Reson. Med.* 2021, *86*, 2179–2191. [CrossRef] [PubMed]
- Wood, J.C.; Ghugre, N. Magnetic Resonance Imaging Assessment of Excess Iron in Thalassemia, Sickle Cell Disease and Other Iron Overload Diseases. *Hemoglobin* 2008, 32, 85–96. [CrossRef]
- Augusto, J.B.; Davies, R.H.; Bhuva, A.N.; Knott, K.D.; Seraphim, A.; Alfarih, M.; Lau, C.; Hughes, R.K.; Lopes, L.R.; Shiwani, H.; et al. Diagnosis and Risk Stratification in Hypertrophic Cardiomyopathy Using Machine Learning Wall Thickness Measurement: A Comparison with Human Test-Retest Performance. *Lancet Digit. Health* 2021, 3, e20–e28. [CrossRef]
- He, T.; Gatehouse, P.D.; Kirk, P.; Tanner, M.A.; Smith, G.C.; Keegan, J.; Mohiaddin, R.H.; Pennell, D.J.; Firmin, D.N. Black-Blood T2* Technique for Myocardial Iron Measurement in Thalassemia. *J. Magn. Reson. Imaging* 2007, 25, 1205–1209. [CrossRef]