



Contents lists available at ScienceDirect

# Computer Methods and Programs in Biomedicine

journal homepage: [www.elsevier.com/locate/cmpb](http://www.elsevier.com/locate/cmpb)

## An accurate and time-efficient deep learning-based system for automated segmentation and reporting of cardiac magnetic resonance-detected ischemic scar

Daniele M Papetti<sup>a</sup>, Kirsten Van Abeelen<sup>b</sup>, Rhodri Davies<sup>c,d,e</sup>, Roberto Menè<sup>b,f</sup>,  
 Francesca Heilbron<sup>f</sup>, Francesco P Perelli<sup>b,f</sup>, Jessica Artico<sup>c,d</sup>, Andreas Seraphim<sup>c,g</sup>,  
 James C Moon<sup>c,d</sup>, Gianfranco Parati<sup>f</sup>, Hui Xue<sup>h</sup>, Peter Kellman<sup>h</sup>, Luigi P Badano<sup>b,f</sup>,  
 Daniela Besozzi<sup>a,i,\*</sup>, Marco S Nobile<sup>i,j,\*</sup>, Camilla Torlasco<sup>f,\*\*</sup>

<sup>a</sup> Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, Milano 20126, Italy

<sup>b</sup> Department of Medicine and Surgery, University of Milano-Bicocca, Milan 20126, Italy

<sup>c</sup> Institute of Cardiovascular Science, University College London, London WC1E 6DD, UK

<sup>d</sup> Barts Heart Centre, St Bartholomew's Hospital, West Smithfield, London EC1A 7BE, UK

<sup>e</sup> MRC Unit for Lifelong Health and Ageing, University College London, London WC1E 6DD, UK

<sup>f</sup> Department of Cardiology, IRCCS Istituto Auxologico Italiano, Via Magnasco 2, Milan 20145, Italy

<sup>g</sup> Department of Cardiac Electrophysiology, Barts Heart Centre, Barts Health NHS Trust, London EC1A 7BE, UK

<sup>h</sup> National Heart, Lung, and Blood Institute, National Institutes of Health, DHHS, Bethesda, MD, USA

<sup>i</sup> Bicocca Bioinformatics Biostatistics and Bioimaging Centre (B4), University of Milano-Bicocca, Veduggio al Lambro 20854, Italy

<sup>j</sup> Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, Mestre, Venice 30172, Italy

### ARTICLE INFO

#### Article history:

Received 22 August 2022

Revised 21 November 2022

Accepted 19 December 2022

#### Keywords:

Deep learning

Cardiovascular imaging

Cardiac magnetic resonance

Myocardial ischemic scar

Image processing

### ABSTRACT

**Background and objectives:** Myocardial infarction scar (MIS) assessment by cardiac magnetic resonance provides prognostic information and guides patients' clinical management. However, MIS segmentation is time-consuming and not performed routinely. This study presents a deep-learning-based computational workflow for the segmentation of left ventricular (LV) MIS, for the first time performed on state-of-the-art dark-blood late gadolinium enhancement (DB-LGE) images, and the computation of MIS transmural extent.

**Methods:** DB-LGE short-axis images of consecutive patients with myocardial infarction were acquired at 1.5T in two centres between Jan 1, 2019, and June 1, 2021. Two convolutional neural network (CNN) models based on the U-Net architecture were trained to sequentially segment the LV and MIS, by processing an incoming series of DB-LGE images. A 5-fold cross-validation was performed to assess the performance of the models. Model outputs were compared respectively with manual (LV endo- and epicardial border) and semi-automated (MIS, 4-Standard Deviation technique) ground truth to assess the accuracy of the segmentation. An automated post-processing and reporting tool was developed, computing MIS extent (expressed as relative infarcted mass) and transmural extent.

**Results:** The dataset included 1355 DB-LGE short-axis images from 144 patients (MIS in 942 images). High performance ( $> 0.85$ ) as measured by the Intersection over Union metric was obtained for both the LV and MIS segmentations on the training sets. The performance for both LV and MIS segmentations was 0.83 on the test sets. Compared to the 4-Standard Deviation segmentation technique, our system was five times quicker ( $< 1$  min versus  $7 \pm 3$  min), and required minimal user interaction.

**Conclusions:** Our solution successfully addresses different issues related to automatic MIS segmentation, including accuracy, time-effectiveness, and the automatic generation of a clinical report.

© 2022 Elsevier B.V. All rights reserved.

**Abbreviations:** BB, bright-blood; BB-LGE, bright-blood late gadolinium enhancement; b-SSFP, balanced-steady-state free precession; CNN, convolution neural network; CMR, cardiovascular magnetic resonance; CV, cross validation; DB-LGE, dark-blood late gadolinium enhancement; DL, deep learning; ICC, intra class correlation; IOU, intersection over union; LV, left ventricle; MIS, myocardial infarction scar; n-SD, n-standard deviation; PSIR, phase-sensitive inversion recovery; ReLU, rectified linear unit; RIM, relative infarcted mass; ROI, region of interest; TI, inversion time.

\* Corresponding authors.

\*\* Corresponding author at: Department of Cardiology, IRCCS Istituto Auxologico Italiano, Via Magnasco 2, Milan 20145, Italy.

E-mail addresses: [hui.xue@nih.gov](mailto:hui.xue@nih.gov) (H. Xue), [daniela.besozzi@unimib.it](mailto:daniela.besozzi@unimib.it) (D. Besozzi), [marco.nobile@unive.it](mailto:marco.nobile@unive.it) (M.S. Nobile), [c.torlasco@auxologico.it](mailto:c.torlasco@auxologico.it) (C. Torlasco).

## 1. Introduction

Cardiovascular magnetic resonance (CMR) is the most advanced non-invasive cardiac imaging test, providing a broad range of information on the heart, and is considered the gold standard for chambers' size, function, and non-invasive tissue characterization. A challenge with CMR is delivery - making it cheaper, more standardized, and faster while retaining accuracy. Solutions will need to address all parts of the imaging chain, including image acquisition, image segmentation (analysis), and reporting processes [1–3].

Deep Learning (DL) approaches have been successfully employed for CMR segmentation tasks, tackling multiple issues simultaneously - accuracy, reproducibility, time effectiveness, and automatic reporting. Examples include ventricular volumes, mass, wall thickness segmentation, and left ventricle (LV) landmarking, where DL-based systems have shown similar accuracy and better reproducibility than human experts [4–6].

The segmentation of ischemic scar is one CMR analysis task that has not been properly addressed. CMR late-gadolinium enhancement (LGE) imaging accurately detects the location, size, and extent of myocardial infarction scar (MIS) [7,8]. These data are linked with prognosis, and provide information about myocardial viability used to guide revascularization. Overall, scar burden correlates with arrhythmia risk [9–12], and the likelihood of segmental functional recovery has an inverse relationship with scar transmural extent (i.e., the extension of the scar in the myocardial wall thickness), with a significantly lower likelihood of benefit from the revascularization in those segments with transmural extent of MIS higher than 50% [13]. MIS may also be used for drug development as a primary endpoint in clinical trials of cardioprotective medication, an area of need [14].

An “optimal” method for CMR LGE quantification has not been identified yet [15]. Both visual assessment and manual contouring lack reproducibility and accuracy against histology [16]. Semi-automated techniques, e.g., the “full-width-half-maximum” or the “ $n$ -Standard Deviation” ( $n$ -SD) techniques, may produce widely differing results for fibrosis quantification but, given adequate standardization, they can improve reproducibility [16–18]. Still, these analyzes are time-consuming, operator-dependent and require training and quality assurance processes. DL-based solutions for MIS segmentation can provide high performances and full automatization. Convolutional Neural Networks (CNNs) with a U-Net architecture [19], which are particularly suited for segmentation tasks in medical imaging thanks to their ability to perform fast and precise segmentation using a reduced amount of learning data, have been successfully applied to the task [20,21]. Examples include MIS segmentation performed independently [22], matched with the LV segmentation on short-axis LGE images [23,24], or in a cascaded multi-view U-Net [25]. The generalizability of these studies is limited though, because they are often monocentric and based on small, private datasets. Furthermore, they are all performed on conventional bright-blood (BB) LGE images, where low contrast between blood pool and scar may determine variability on LV boundaries annotation, and increased difficulty in identifying thin, subendocardial MIS [26]. Alternative heuristics previously exploited for LGE segmentation include thresholding (e.g., using Otsu's algorithm) and clustering [17].

We here for the first time explore the use of dark-blood (DB) LGE, the current gold standard sequence for the detection of subendocardial scar [26], combined with two CNNs with U-Net architecture, to perform an automated, time-efficient and accurate MIS segmentation [27].

## 2. Methods

### 2.1. Ethics approval and informed consent

Patients' images were collected retrospectively at IRCCS Istituto Auxologico Italiano (Milan, Italy) and Barts Heart Center (London, UK). The local Ethics Committees of both centres approved the study. All patients provided written consent for anonymized use of data for research purposes at the time of the CMR.

### 2.2. Imaging and MRI acquisition

State-of-the-art free-breathing, motion-corrected, dark blood (DB) phase-sensitive inversion recovery (PSIR) LGE short-axis images were retrospectively collected on three 1.5T scanners (1x AvantoFit, 2x Aera; Siemens, Erlangen, Germany), from 1 January 2019 to 1 June 2021, as previously described [26]. LV short-axis stacks with at least one slice displaying subendocardial/transmural scar were included. Images of non-diagnostic quality due to any reason and stacks showing only non-ischemic LGE patterns were excluded. Prior to analysis, images were entirely anonymized.

### 2.3. Ground truth labeling and image pre-processing

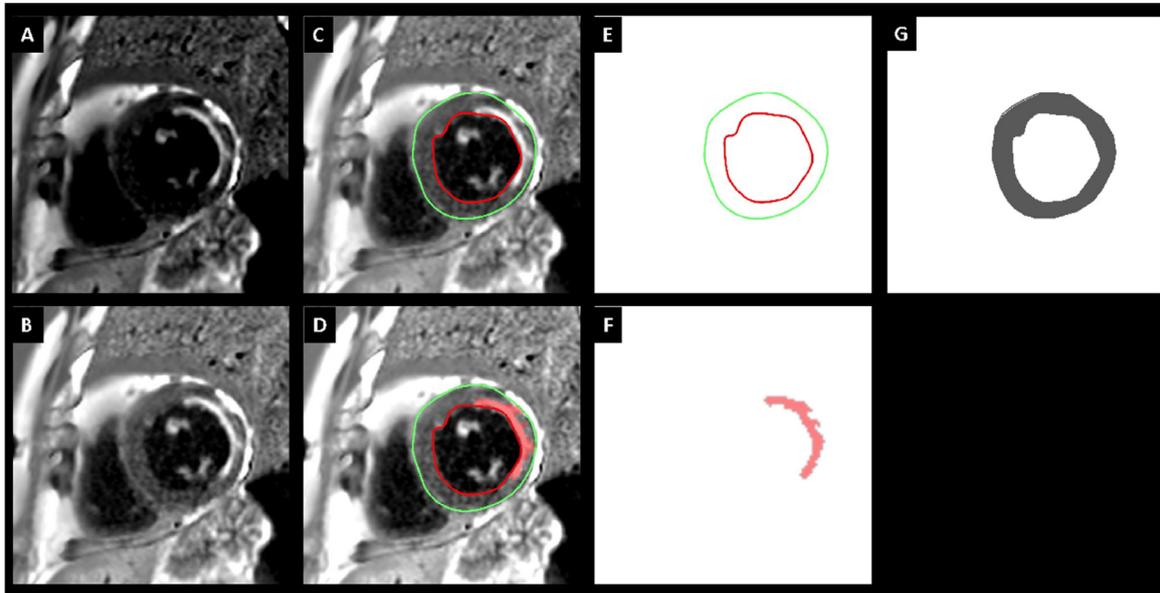
LV and MIS segmentations were performed by IRCCS Istituto Auxologico Italiano investigators with a dedicated open-source software (Horos Purview, Annapolis, USA). Two issues had to be addressed: DB-LGE images windowing, and the MIS segmentation technique. With regards to windowing, a level range (WL:  $4100 \pm 100$ ; WW:  $1300 \pm 50$ ; see Fig. 1, panel A and panel B) was chosen visually, to provide an optimal visualization of the myocardium and the blood pool. The choice was corroborated in a sample of 6 patients (56 slices), in whom LV mass calculated on windowed DB-LGE images was compared to LV mass calculated on matching bright-blood Magnitude-IR LGE (BB-LGE) images and balanced-steady-state free precession (b-SSFP) cine images by two independent observers (manual LV endo- and epicardial borders segmentation, mass calculation by Simpson's disk stack technique).

To choose a scar segmentation technique, MIS was assessed on the same set of 56 DB-LGE slices with different techniques (2-SD, 3-SD, 4-SD, 5-SD, 6-SD); the results were then compared against MIS segmented on matching BB-LGE images with the 5-SD technique as the gold standard [28]. Interobserver reproducibility was evaluated on DB-LGE MIS segmentation (all techniques) by Intra Class Correlation (ICC). Normality of data was tested with the Kolmogorov-Smirnov test of normality.

Eventually, LV endo- and epicardial borders were manually traced on windowed images on the whole dataset, and the 4-SD technique was applied to segment MIS. Briefly, a region of interest (ROI) was selected in the remote myocardium; then, the mean and the SD of the selected ROI were calculated, and pixels with an intensity greater than 4-SD above the mean were defined as scarred [16]. Fig. 1 details the labeling and mask creation process; Fig. S1 details examples of different  $n$ -SD segmentation.

All images were exported in TIFF format (300dpi), and the central field of view ( $512 \times 512$  pixels) was cropped in each. A 2-fold data augmentation process was performed by means of contrast stretching (i.e., as generated by the magnetic resonance scanner and windowed, as described above - see Fig. 1, panels A and B).

The area between the endo- and epicardial border, i.e. the LV myocardium, was converted into a shape to create the LV mask for the CNN.



**Fig. 1.** Labelling and mask creation example. The original image (panel A) was windowed for a better definition of the heart wall (WL:  $4100 \pm 100$ ; WW:  $1300 \pm 50$ ) (panel B). Endo- and epicardial borders (respectively, red and green contours) were manually traced on the windowed image; papillary muscles were included in the blood volume (panel C). The ischemic scar was segmented with the 4-Standard Deviation technique, with a region of interest (ROI) drawn in the remote myocardium (panel D). The remote ROI was selected to be as close as possible to the scar while avoiding the insertion points, the basal septum, and any area of non-ischemic late gadolinium enhancement. A mask for myocardium (panel E) and a mask for scar segmentation (panel F) were generated. The myocardium mask was then converted into a shape (panel G) prior to being fed to the CNNs.

#### 2.4. CNN models and training

We trained two CNNs based on the same U-Net architecture. The CNNs were developed using Python 3.7.11, NumPy 1.21.2 and PyTorch 1.10 with CUDA 11.3 [29,30]. The U-Net architecture consists of a set of contraction blocks followed by a set of expanding blocks, whereby the size of the input image is first reduced while extracting the hidden features information; then, these high-resolution features are combined again with the spatial information to reconstruct an output image that represents the ROI of the input image. In this work, contraction blocks consist of a max pooling operator followed by two 2-dimensional convolutional layers. Each convolutional layer is provided with an activation function and a regularization operator (see Supplementary Material, Table S1). A batch normalization layer is also used to reduce overfitting [29]. Expanding blocks consist of a transpose convolutional layer followed by two convolutional layers, as reported in Supplementary Material, Table S2. Each expanding block receives two inputs: the first one is a tensor generated by the previous block (whichever the type) and fed into the transpose convolutional layer; the second one is an output of a contraction block that is concatenated with the output of the transpose convolution along with the filters, and fed into the first convolutional layer.

The overall topology of the CNNs used in this work is shown in Fig. 2. Before and after the building blocks of the U-Nets, two 2-dimensional convolutional layers, activated by ReLU functions, are employed (see Supplementary Material, Table S3). For a layer-wise description of the architecture, we refer to Supplementary Material, Table S4.

The CNNs receive as input a  $512 \times 512$  gray shaded image, whose pixels are normalized in the  $[0, 1]$  interval, and provide as output a  $512 \times 512$  matrix of real values in the  $[0, 1]$  interval that can be converted into a  $512 \times 512$  gray shaded image (see Fig. 3, panel A). The pixels of the output images are gray shaded: each dark pixel represents the respective pixel in the input image belonging to a ROI, that is, an LV region or scar tissue (see Fig. 3, panels B and C). To compute all the performance metrics consid-

ered in this work, the values of the output matrix are rounded to produce binary values.

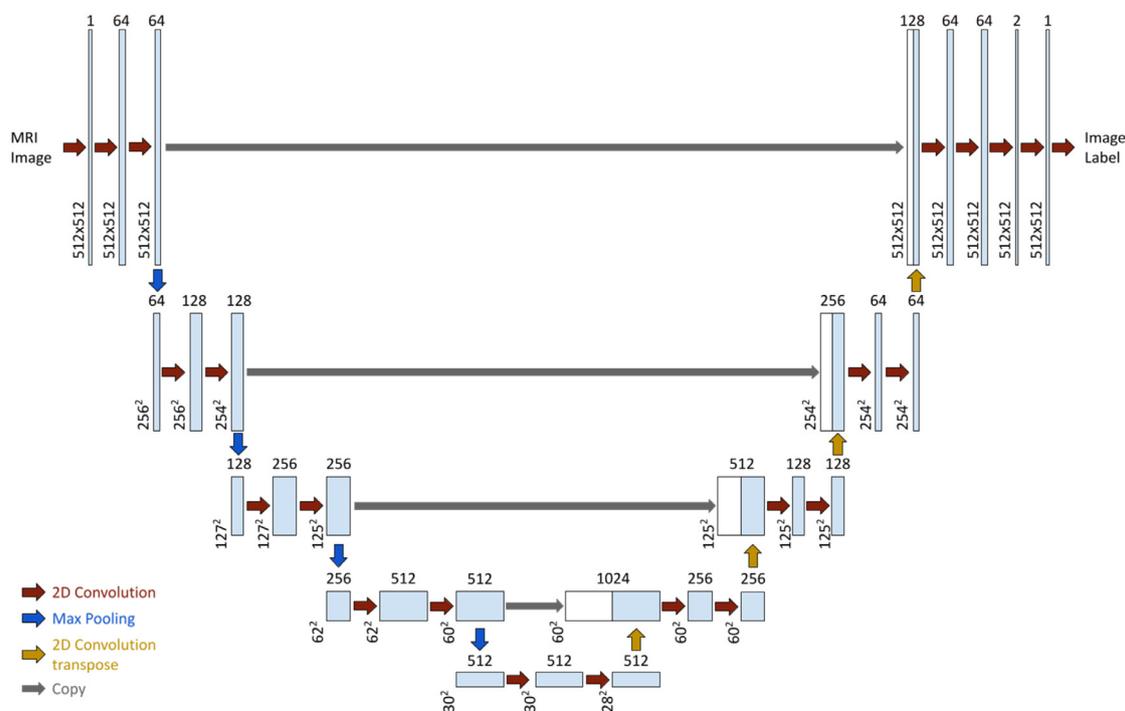
The weights of the CNNs were initialized according to PyTorch's truncated normal distribution with mean equal to 0 and standard deviation equal to  $\sqrt{\frac{2}{fan_{in}}}$ , where  $fan_{in}$  is the number of inputs to the layer [31]. Both the model to detect the LV region and the model to detect the MIS were trained for 100 epochs, and an early stopping criterion was applied if the loss function did not improve of at least  $5 \cdot 10^{-4}$  for 5 epochs in a row. In the case the early stopping criterion was satisfied, the training process was interrupted and the model was restored to the best model found, i.e., the weights were reset to the weights of the last epoch when the loss function improved. During each epoch, prior to feeding the images of the training set to the CNNs, such images were randomly perturbed by means of shift, rotation, and zoom operators. The shift operator shifts both horizontally and vertically the image of a maximum of 26 pixels, the rotation operator rotates the image to a maximum of  $5^\circ$  either in a clockwise or counterclockwise direction, while the zoom operator can zoom the image either in or out of a 10% scale.

#### 2.5. Performance analysis

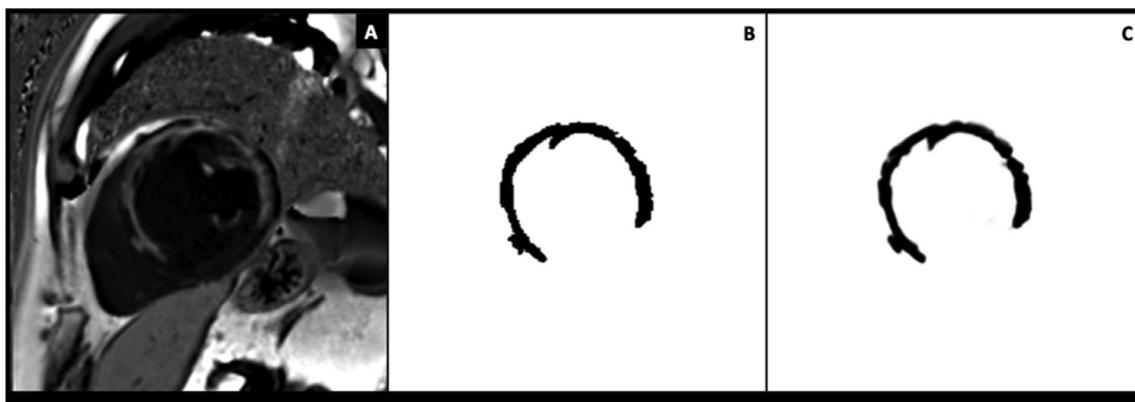
The dataset was split into a training set and a test set according to a 90-10 policy, that is, 90% of the images were used to train and validate the CNN models and the remaining 10% to monitor their generalization capability as a test set. The performance of the CNNs was assessed by means of a 5-fold cross-validation (CV) process performed by leveraging the training set. Six performance metrics were evaluated during the training: binary cross-entropy (loss function), accuracy, precision, recall, Intersection over Union (IoU), and Dice score.

The binary cross-entropy is defined as:

$$l_k^{BCE} = -\frac{1}{K} \sum_{k=0}^K [y_k \cdot \log(x_k) + (1 - y_k) \cdot \log(1 - x_k)],$$



**Fig. 2.** Topology of the Convolutional Neural Networks (U-Net architecture) employed in this work for automatic MIS segmentation. Contraction blocks (on the left) reduce the size of the input image and extract multiple hidden features represented in the different layers. Expanding blocks (on the right) reconstruct an image representing the ROI of the input image by composing the extracted hidden features. The convolutional layers of contracting and expanding blocks are based on a  $3 \times 3$  kernel with padding equal to 1. In this type of architecture, the output of a contraction block is propagated as input to the next contraction block, and as one of the two inputs of the respective expanding block. The output of the transpose convolution is padded to match the same width and height of the output of the respective contraction block. The resulting tensor and the second input tensor are concatenated along with the filters. The resulting tensor is fed into the first convolutional layer of the expanding block. In our architecture, the last contraction block does not double the number of filters.



**Fig. 3.** Scar segmentation prediction example. Panel A: original image. Panel B: scar mask obtained using the 4-Standard Deviation segmentation technique. Panel C: Deep Learning prediction.

where  $K$  is the number of images in the batch,  $x_k$  is the prediction of the  $k^{\text{th}}$  image, and  $y_k$  is the respective ground truth. Since both  $x_k$  and  $y_k$  are images, pixel-wise operations were performed to compute the loss function. The accuracy, precision and recall scores are evaluated on the prediction of LV or MIS, according to the CNN task.

The IoU evaluates the ratio between the intersection of the ground truth and the predicted ROI over the union of such regions (see Supplementary Material, Fig. S2). The ROI is the set of pixels that share the same label value, e.g., all pixels that are predicted as 0. In this case, the ROIs are the pixels predicted either as LV or MIS, according to the CNN model. Similarly, the Dice score computes the ratio between the intersection of the ground truth and the predicted ROI over the combined number of pixels classified as a ROI in both images.

The precision, recall and Dice scores were computed only on the images presenting a LV or a MIS. At the end of the training process with the whole dataset, we labelled the LV slices from 30 patients from the test set as “basal”, “mid-ventricular” and “apical”, and we evaluated the performance of the two CNNs in the different LV locations.

### 2.6. Clinical report

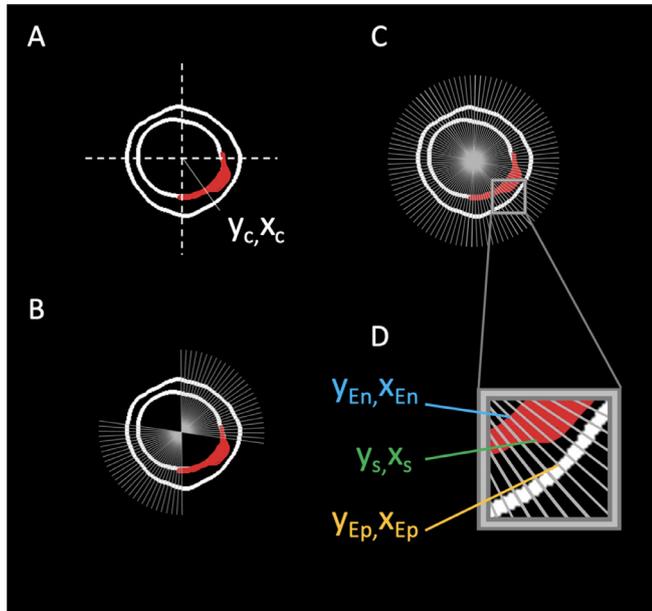
The resulting DL segmentations were processed using Python 3.8.5 and OpenCV 4.5.5.62 to generate a clinical report. This report consisted of two metrics describing MIS extent, both with demonstrated prognostic implications [9–12]:

- (1) Relative infarcted mass (RIM), i.e., MIS mass as a percentage of the LV mass, calculated on the DB-LGE images.

**Table 1**

Interobserver variability. LV mass and MIS quantification performed using different techniques by two independent operators on dark-blood late gadolinium enhancement images (56 slices, 6 patients). Interobserver variability is measured by Intra Class Correlation (ICC). Data are expressed as mean (SD). LV: left ventricle; SD: standard deviation.

N = 56 slices	LV Mass	2-SD	3-SD	4-SD	5-SD	6-SD
Observer 1, g	13.1 (2.2)	5.2 (2.2)	3.8 (1.7)	3.1 (1.7)	2.5 (1.6)	2.3 (1.3)
Observer 2, g	13.2 (2.0)	5.2 (2.0)	3.9 (1.9)	3.0 (1.8)	2.6 (1.5)	2.3 (1.3)
ICC	0.94	0.90	0.88	0.95	0.96	0.95



**Fig. 4.** Transmurality assessment. The centroid of the contour is given by the coordinates  $(x_c, y_c)$  (panel A). From there, the left ventricle is divided into 360 chords (panels B and C). In the presence of MIS (panel C and D), the coordinates of the intersection points between the endo- and epicardial contour are used to calculate the LV wall thickness (panel D, respectively  $x_{en}, y_{en}$  and  $x_{ep}, y_{ep}$ ), and the coordinates of the intersection points between the segmented scar borders ( $x_s, y_s$ ) are used to calculate MIS thickness. Transmurality percentage is then calculated as (MIS thickness/LV wall thickness) \* 100 for each cord intersecting the scar, resulting in an average transmurality percentage for the MIS.

- (2) Transmurality. Each LV slice was segmented into 360 chords where, for each chord, the transmural extent of MIS - that is, the percentage of the wall thickness which is infarcted - was computed (see Fig. 4) [32]. When present, the average MIS transmural extent was categorized as <25%, 25–50%, 50–75% or >75%, and the distribution of MIS, expressed as the percentage of the LV over the four categories and averaged over the number of LV slices, was reported.

### 3. Results

#### 3.1. Images dataset

A total of 2704 DB-LGE anonymized images were collected from 441 patients. Among those, 1454 images from 297 patients were excluded (quality issues in 74 images; see Supplementary Material, Fig. S3, for details). The final sample consisted of 1355 short-axis DB-LGE images from 144 patients; MIS was present in 942 images, 100% of patients. On average, 4-SD semi-automated analysis of a 10-slices short-axis stack took  $7 \pm 3$  min.

#### 3.2. Segmentation technique and interobserver reproducibility

LV mass and MIS were normally distributed, irrespectively of technique and sequence.

**Table 2**

LV mass measured on different sequences on 6 patients. Data are expressed as mean (SD). Significance is tested by one-way ANOVA for repeated measures. b-SFFP cine: balanced-steady-state free precession cine; DB: dark blood; BB: bright blood; LGE: late gadolinium enhancement.

N = 6 patients	b-SFFP cine	DB-LGE	BB-LGE	p
Mean mass, g	116.7 (31.6)	120.4 (34.2)	121.6 (38.4)	0.33

Interobserver variability was good for LV mass calculation and for all MIS segmentation techniques on DB-LGE images (Table 1).

LV mass did not differ significantly when calculated from DB-LGE (WL:  $4100 \pm 100$ ; WW:  $1300 \pm 50$ ), BB-LGE and cine images (Table 2).

Mean MIS volume increased as the threshold varied from 6- to 2-SD from mean remote myocardium signal (see Supplementary Material, Figs. S1 and S3). Among all the tested segmentation techniques, DB 4-SD showed the best agreement by the ICC when compared to the gold standard BB 5-SD (Table 3).

#### 3.3. Performance of the convolutional neural networks

In all tests for LV and MIS segmentation, the training processes were performed by using an NVIDIA RTX 3090. While each training required more than 2 h and 30 min, the time required for predicting the segmentation of a 10-slices short-axis is approximately 8 s.

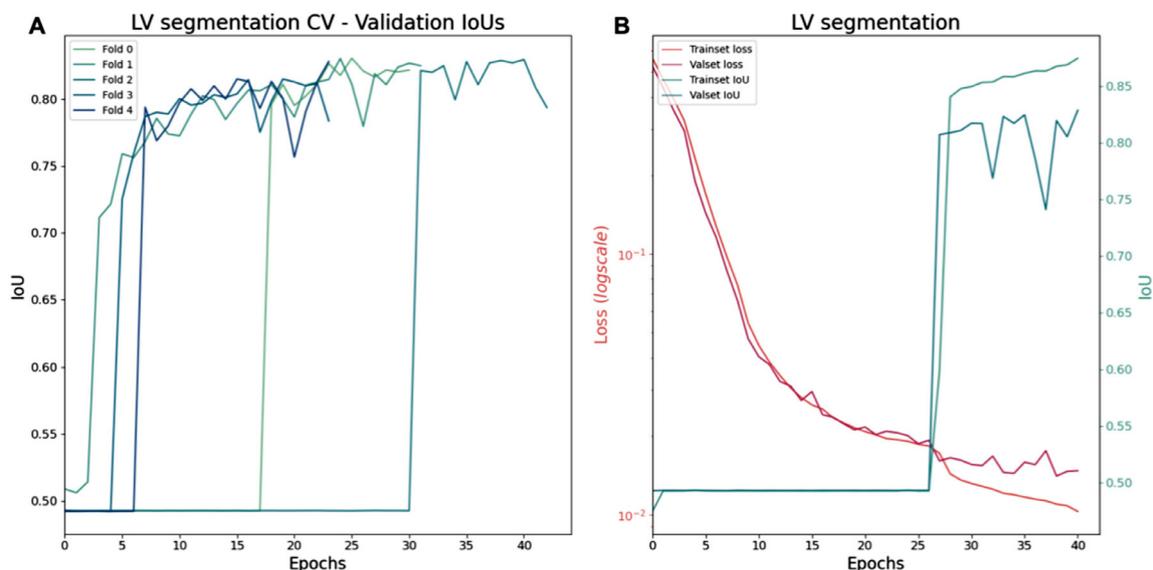
##### 3.3.1. LV segmentation CNN

We adopted a 5-fold CV, online data augmentation and an early stopping criterion for the training of both CNNs to prevent overfitting. This strategy was successful in all 5 iterations, with an IoU around 0.82 (see Fig. 5A), an accuracy of 0.99 and a Dice score around 0.73 (see Supplementary Material, Table S5). During the training, the loss decreased by an order of magnitude after a few epochs, as shown in Supplementary Material, Fig. S4 (panel A). The performance obtained by the networks on the held-out fold (i.e., the validation set) shows that the CNNs did not overfit, as reported in Supplementary Material, Table S5. After assessing the performance of the CNN for the LV segmentation task, we trained the CNN using the whole training set. The training set was split according to a 90–10 policy: 90% assigned to the training, and 10% assigned to the validation set to monitor the generalization capability of the models. We performed online data augmentation

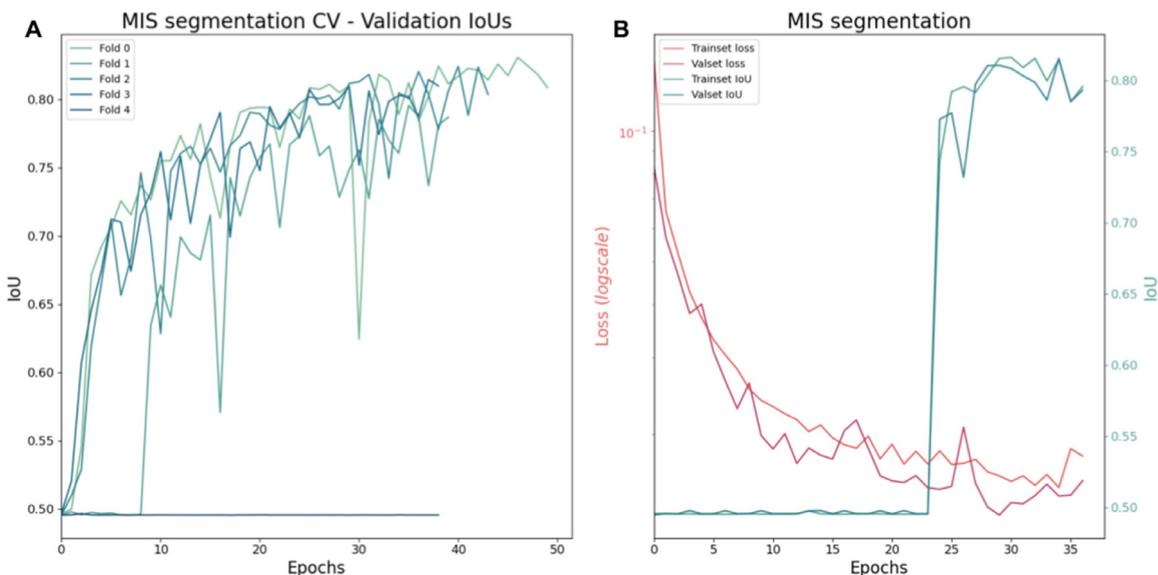
**Table 3**

Segmentation techniques comparison. Each technique of MIS segmentation on DB-LGE is compared against the gold standard 5-SD BB-LGE. Data are reported as mean (SD). MIS: myocardial infarction scar; BB: bright blood; DB: dark blood; ICC: Intra Class Correlation; SD: standard deviation.

	MIS/slice	MIS/patient	ICC
BB 5-SD g	3.1 (1.9)	24.9 (12.8)	
DB 2-SD g	5.1 (2.1)	45.3 (14.8)	0.47
DB 3-SD g	3.6 (1.8)	32.9 (11.6)	0.86
DB 4-SD g	2.9 (1.6)	25.9 (12.0)	0.94
DB 5-SD g	2.4 (1.4)	20.8 (10.7)	0.89
DB 6-SD g	2.1 (1.6)	18.3 (10.2)	0.77



**Fig. 5.** IoU (panel A) computed on the validation set during the 5-fold CV process for LV segmentation. Loss and IoU (panel B) computed on both the training and validation sets during the training of the CNNs for LV segmentation. The y axis is log-scaled for the loss function.



**Fig. 6.** IoU (panel A) computed on the validation set during the 5-fold CV process for LV segmentation. Loss and IoU (panel B) computed on both the training and validation sets during the training of the CNNs for MIS segmentation. The y axis is log-scaled for the loss function.

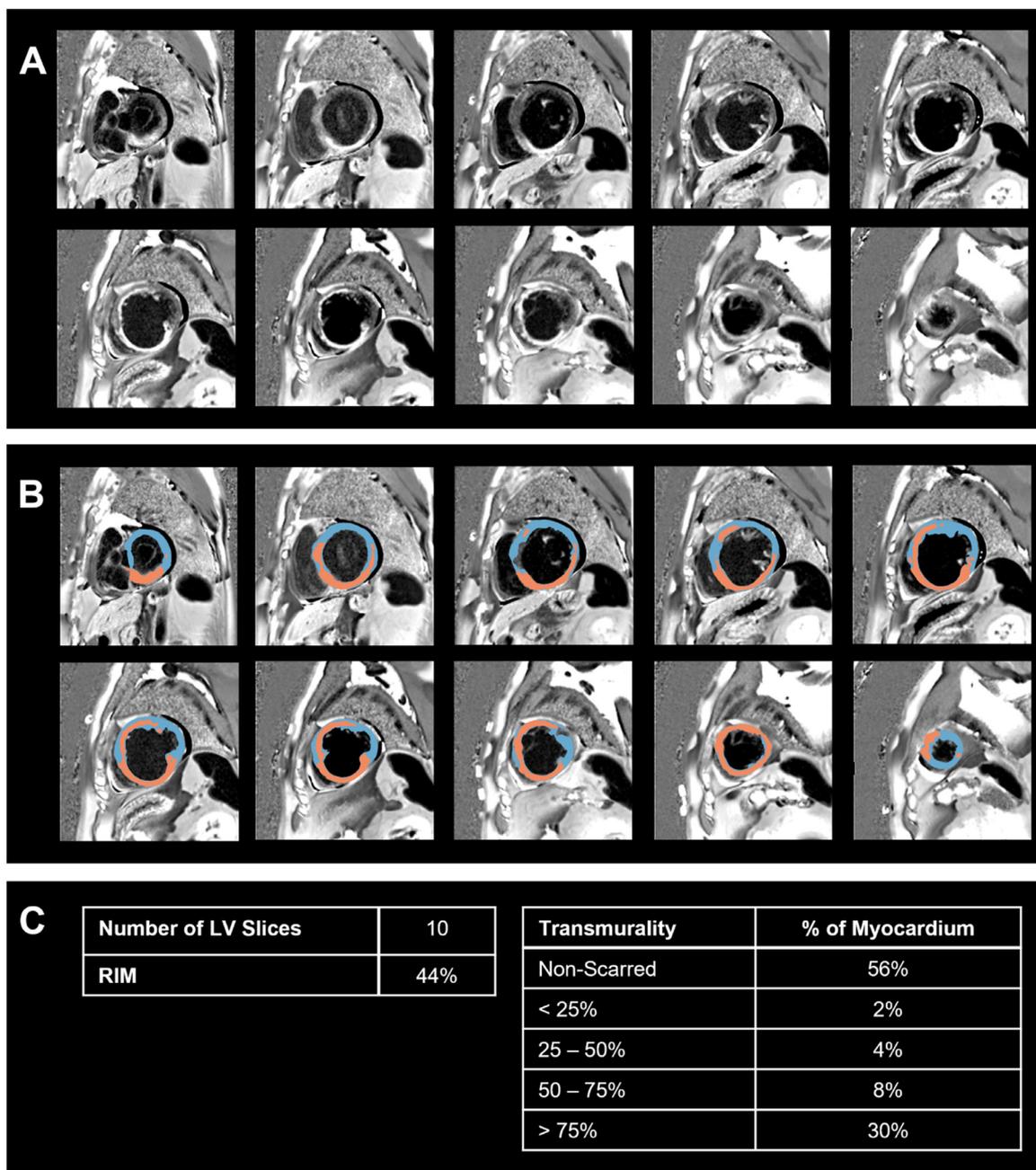
as described above, which was applied during each epoch of the training. Fig. 5B shows the performance metrics computed for each epoch during the training process. The results confirm what was observed with the 5-fold CV: the loss quickly decreases by an order of magnitude, while the IoU computed on the validation set improves to 0.84. At the end of the training process, the CNN was evaluated on test set images. The performance in terms of IoU, accuracy and Dice score (see Supplementary Material, Table S6) are similar to the performance obtained with the 5-fold CV, with an IoU equal to 0.84, thus confirming the generalization capability of the models.

Concerning the performance in different LV locations, the CNN model is slightly better on the mid-cavity images with an IoU of 0.88 against an IoU of 0.85 on apical and basal images.

### 3.2.2. MIS segmentation CNN

The evolution of the performance metrics computed on the training sets during the 5-fold CV process is shown in Fig. 6. Data

augmentation and early stopping criterion were applied as in the LV segmentation CNN. This strategy was successful in 4 out of 5 iterations, with an IoU around 0.8 (see Fig. 6A), an accuracy of 0.99 and a Dice score of 0.76 (see Supplementary Material, Table S7). During these 4 iterations of the CV, the loss dropped to an order of magnitude lower after a few epochs, as shown in Supplementary Material, Fig. S3 (panel B). Such a decrease does not immediately yield a better performance in terms of IoU. As shown in Fig. 6A, the CNNs required several epochs to improve the segmentation quality of the image, reaching an IoU around 0.8 on the validation set. The metrics computed on the validation sets during the 5-fold CV are reported in Supplementary Material, Table S7. Finally, Fig. 6B shows the performance of the CNN for MIS segmentation using the whole training set. During the training, the loss slightly decreased while the IoU computed on the validation set improved to 0.8. Also in this case, the training set was split according to the 90–10 policy, and the validation set was used to monitor the generalization performance of the CNN during the training process. The trained CNN



**Fig. 7.** Reporting output (I). Panel A: pre-processed late-gadolinium enhancement (LGE) images; Panel B: Deep Learning-predicted segmentations (post-processed) for the left ventricle (LV) (blue) and the myocardial infarction scar (MIS) (red). Panel C: transmurality metrics: the number of considered LV slices, the RIM, and transmurality distribution of MIS over the slices, are reported. RIM: Relative Infarcted Mass, i.e. LGE mass as a percentage of the LV mass.

was used to predict the MIS in the test set images. Also in this case, the performance on the test set (see Supplementary Material, Table S6), with an IoU equal to 0.83, confirms the generalization capability of the CNN.

The performance of the CNN was evaluated on different locations of the heart for 30 patients. The results show that the CNN equally performs for basal and mid-cavity regions of the LV with an IoU around 0.79, while an IoU of 0.72 is obtained on the apical regions.

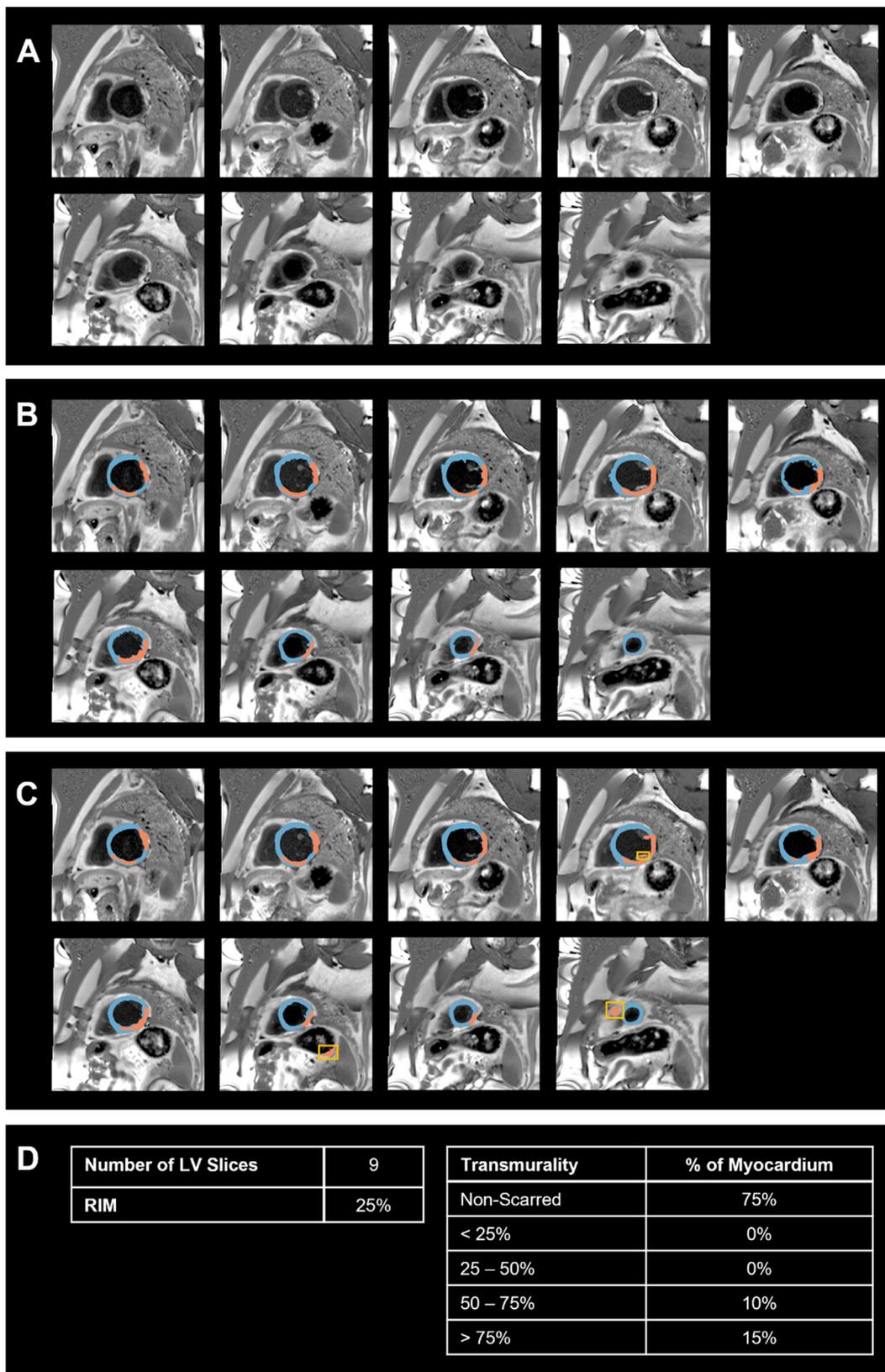
### 3.4. Reporting

The generation of the clinical report took, on average, 12 s (AMD Ryzen 5800H, 16GB RAM, NVIDIA RTX 3060) for a 10-slices short-axis stack. The report displayed the original DB-LGE images

fed to the CNNs, the DL-segmentation overlapped on the original images, and a summary table of RIM and average MIS transmurality (Fig. 7). Inaccuracies were addressed through simple post-processing (OpenCV 4.5.5.62) of the predictions, where noise in the LV segmentation (retaining the largest segmentation) and scar segmentations (only including segmentations that overlap with the LV segmentation) outside the LV were automatically removed (Fig. 8). Visual comparison of the predictions against the original image allowed immediate quality control by the clinician.

### 4. Discussion

In this study, we developed a CNN-based workflow for semiautomatic MIS segmentation on DB-LGE CMR images, demonstrated its accuracy, and proposed an *ad hoc* clinical reporting output.



**Fig. 8.** Reporting output (II). Panel A: pre-processed late-gadolinium enhancement (LGE) images. Panel B: ground-truth labels for the left ventricle (LV) (blue) and the myocardial infarction scar (MIS) (red). Panel C: Deep Learning-predicted LV and MIS segmentations; from left to right, top to bottom, MIS segmentation inaccuracies (yellow bounding boxes) can be found in the fourth (papillary muscles segmented as MIS), seventh (a digestive tract region segmented as MIS), and ninth image (right ventricle segmented as MIS). These were removed through manual post-processing. Panel D: reporting metrics as described in Fig. 7. RIM: Relative Infarcted Mass, i.e. LGE mass as a percentage of the LV mass.

The proposed solution has several strengths. First, to the best of our knowledge, this is the first attempt to perform DL-based segmentation on state-of-the-art DB-LGE imaging, which optimizes the contrast between blood pool and subendocardial LGE, and thus guarantees the maximum sensitivity in the detection of MIS [33].

Due to the novelty of DB-LGE imaging, there is no standardized method to perform the LV/MIS segmentation tasks. Windowing strongly affects subendocardial boundaries contouring, which in turn strongly affects MIS size. Here, we selected a windowing range that looked visually satisfactory to distinguish LV from blood pool, and we corroborated this threshold by comparing LV mass calculated on windowed DB-LGE against cine images.

Regarding MIS segmentation, in a recent animal study by Nies et al. [28] 5-SD was found to be the most accurate semi-automated technique against histology to segment MIS both on standard BB-PSIR and in PSIR with TI set to nullify the blood, i.e., with “dark blood”. In our sample, 4-SD DB-LGE segmentation was the most consistent with 5-SD BB-LGE segmentation, with an excellent inter-observer reproducibility for all  $n$ -SD DB-LGE techniques with  $n \geq 4$ . We argue that this discrepancy may depend on the different pulse sequence used, and on the small sample size (24 LGE slices by Nies et al., 56 LGE slices in the present study). Indeed, also on visual assessment, the 5-SD technique seems to miss some thin and faded subendocardial MIS on DB-LGE images (see Supplementary Material, Fig. S1) [15].

With these caveats, our system performs accurately against ground truth, and requires 80% less computational time and minimal user interaction compared to  $n$ -SD techniques. Indeed, the DL computational time was lower than 1 min/patient for segmentation and clinical report elaboration versus  $\sim 5$ –10 min/patient needed by a doctor to perform segmentation only. Image cropping and, only if needed, manual correction of scar segmentation output were the only inputs required by the operator. Thus, our proposed solution overcomes the limitations of both visual assessment and semiautomatic techniques: the poor reproducibility of visual assessment, as well as the time needed for the manual LV segmentation and delineation of ROIs for threshold computation, are indeed the main factors determining, respectively, reporting inaccuracies and under-usage of quantitative tools in clinical routine. Our solution has significant potential for clinical usage, already offering an automatic clinical reporting tool. Such a report consists of a table that summarizes all the findings. The display of DL-based LV and MIS segmentation results allows immediate visual checking of accuracy (as shown in Figs. 7 and 8) and the localization of the infarcted areas. The table also displays the RIM, and describes the infarction burden by transmural, providing prognostic information and data to guide management, and a suitable end-point for clinical trials [34].

From a development perspective, our solution is based on the U-Net architecture, which is extensively adopted in the medical imaging field thanks to its ability to successfully perform image segmentation in the presence of highly noisy images. The medium-sized dataset we used, consisting of 1355 images, represents a compromise between the need to provide an adequate number of examples for the CNNs to learn and the time required to perform image labeling (in this study, approximately 24 person-hours of trained operators). Performing a 2-fold data augmentation by contrast stretching on the whole dataset, the dataset was expanded to 2710 images. Such a dataset was split into a training set and test set according to a 90–10 policy. A different training set was obtained for each training epoch by applying shift, rotation and zoom operators. U-Nets for myocardial scar segmentation were previously exploited [24,32] but, differently from our work, the authors were aiming at volumetric visualization and did not exploit DB-LGE,

which does not allow to carry out a direct comparison of the performance of the two DL frameworks. To the best of our knowledge, our work represents the first attempt at segmenting human DB-LGE images using DL. Alternative heuristics for LGE segmentations include thresholding (e.g., using Otsu’s algorithm) and clustering [32].

During the 5-fold CV process used to evaluate the CNNs, we encountered a potential issue regarding the weights’ initialization. For some folds, the CNNs started with a greater loss and required a higher number of epochs to converge - or they did not converge at all, as in the case of the LV segmentation - possibly because of the stochasticity of weights’ initialization. We handled this issue by early detection and restarting the training with a better initialization. A test set, independent from both the training and validation sets, was used to check whether the trained CNN models were possibly biased towards the training and/or the validation sets. Our results show that both models converged and could generalize the basic structures in the data - without learning specific structures about the data presented to the CNNs - being thus able to perform well on new data samples characterized by a feature space similar to the input data. The CNNs also show a comparable performance in the different locations of the heart (apical, basal and mid-cavity).

The computational training time (about 2.5 h per CNN) was consistent with the optimized architecture.

We are aware of a few current limitations of our study, and plan to solve them soon.

First, as already discussed, further research is needed to standardize LV/MIS segmentation on DB-LGE images. It must be noted, though, that MIS detection is independent of LV detection in our workflow, and that the reporting tool is still a proof-of-concept. It was not possible to test our system on a publicly available dataset, or directly compare it to other systems already available, because this is the first experience on DG-LGE images. Also, training and validation has been performed on the whole dataset, so it is not currently possible to assess the multicenter/multiscanner generalizability. Although we did not train our system to identify the different LV segments, an essential knowledge for any CMR doctor, the available reporting output provides quantitative information with a prognostic value and the potential to guide clinical decision-making. Future developments will include training the LV CNN to identify the 16 segments according to the American Heart Association classification.

Finally, although the computational process is fast, our system currently works as a stand-alone. Images must be exported from the MR scanner and fed to the CNNs, limiting the full integration of our system in current clinical workflows. To overcome this limitation and promote the clinical adoption of the proposed tool, we plan to integrate our methodology into MR scanners, as it already happened for other DL-based solutions [4]. Inline processing would ensure immediately available results, prior to the next image series.

## 5. Conclusion

Our solution successfully addresses different issues related to automatic MIS segmentation, including accuracy, time-effectiveness, and the automatic generation of a clinical report. Further research is needed to standardize segmentation on DB-LGE images.

## Funding

The Italian Ministry of Health partly funded this research.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We would like to acknowledge Vasco Coelho for his help in the execution of the tests required by the Reviewers.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cmpb.2022.107321](https://doi.org/10.1016/j.cmpb.2022.107321).

## References

- [1] K. Menacho-Medina, N.A.B. Ntusi, J.C. Moon, J.M. Walker, R. Jacob, Rapid cardiac MRI protocols: feasibility and potential applications, *Curr. Radiol. Rep.* 8 (2) (2020), doi:[10.1007/s40134-020-0344-6](https://doi.org/10.1007/s40134-020-0344-6).
- [2] C. Torlasco, et al., Effective study: development and application of a question-driven, time-effective cardiac magnetic resonance scanning protocol, *JAHA* 11 (1) (2022) e022605, doi:[10.1161/JAHA.121.022605](https://doi.org/10.1161/JAHA.121.022605).
- [3] B. Ruijsink, et al., Fully automated, quality-controlled cardiac analysis from CMR, *JACC Cardiovasc. Imaging* 13 (3) (2020) 684–695, doi:[10.1016/j.jcmg.2019.05.030](https://doi.org/10.1016/j.jcmg.2019.05.030).
- [4] H. Xue, J. Artico, M. Fontana, J.C. Moon, R.H. Davies, P. Kellman, Landmark detection in cardiac MRI using a convolutional neural network, *Radiol. Artif. Intell.* (2021) e200197, doi:[10.1148/ryai.2021200197](https://doi.org/10.1148/ryai.2021200197).
- [5] H. Xue, et al., Automated inline analysis of myocardial perfusion MRI with deep learning, *Radiol. Artif. Intell.* 2 (6) (2020) e200009, doi:[10.1148/ryai.2020200009](https://doi.org/10.1148/ryai.2020200009).
- [6] S. Ghadimi, et al., Fully-automated global and segmental strain analysis of DENSE cardiovascular magnetic resonance using deep learning for segmentation and phase unwrapping, *J. Cardiovasc. Magn. Reson.* 23 (1) (2021) 20, doi:[10.1186/s12968-021-00712-9](https://doi.org/10.1186/s12968-021-00712-9).
- [7] G.D. Aquaro, et al., Clinical recommendations of cardiac magnetic resonance, part I: ischemic and valvular heart disease, *J. Cardiovasc. Med.* 18 (4) (2017) 197–208, doi:[10.2459/JCM.0000000000000498](https://doi.org/10.2459/JCM.0000000000000498).
- [8] T. Emrich, M. Halfmann, U.J. Schoepf, K.F. Kreitner, CMR for myocardial characterization in ischemic heart disease: state-of-the-art and future developments, *Eur. Radiol. Exp.* 5 (1) (2021) 14, doi:[10.1186/s41747-021-00208-2](https://doi.org/10.1186/s41747-021-00208-2).
- [9] M. Disertori, et al., Myocardial fibrosis assessment by LGE is a powerful predictor of ventricular tachyarrhythmias in ischemic and nonischemic LV dysfunction: A meta-analysis, *JACC Cardiovasc. Imaging* 9 (9) (2016) 1046–1055, doi:[10.1016/j.jcmg.2016.01.033](https://doi.org/10.1016/j.jcmg.2016.01.033).
- [10] B.Y.C. Cheong, et al., Prognostic significance of delayed-enhancement magnetic resonance imaging: survival of 857 patients with and without left ventricular dysfunction, *Circulation* 120 (21) (2009) 2069–2076, doi:[10.1161/CIRCULATIONAHA.109.852517](https://doi.org/10.1161/CIRCULATIONAHA.109.852517).
- [11] P.A. Scott, et al., The extent of left ventricular scar quantified by late gadolinium enhancement MRI is associated with spontaneous ventricular arrhythmias in patients with coronary artery disease and implantable cardioverter-defibrillators, *Circ. Arrhythm. Electrophysiol.* 4 (3) (2011) 324–330, doi:[10.1161/CIRCEP.110.959544](https://doi.org/10.1161/CIRCEP.110.959544).
- [12] P. Gao, et al., Prediction of arrhythmic events in ischemic and dilated cardiomyopathy patients referred for implantable cardiac defibrillator: evaluation of multiple scar quantification measures for late gadolinium enhancement magnetic resonance imaging, *Circ. Cardiovasc. Imaging* 5 (4) (2012) 448–456, doi:[10.1161/CIRCIMAGING.111.971549](https://doi.org/10.1161/CIRCIMAGING.111.971549).
- [13] A.G. Almeida, et al., Multimodality imaging of myocardial viability: an expert consensus document from the European Association of Cardiovascular Imaging (EACVI), *Eur. Heart J. Cardiovasc. Imaging* 22 (8) (2021) e97–e125, doi:[10.1093/ehjci/jeab053](https://doi.org/10.1093/ehjci/jeab053).
- [14] S. Desch, et al., Cardiac magnetic resonance imaging parameters as surrogate endpoints in clinical trials of acute myocardial infarction, *Trials* 12 (1) (2011) 204, doi:[10.1186/1745-6215-12-204](https://doi.org/10.1186/1745-6215-12-204).
- [15] J. Schulz-Menger, et al., Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update: Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing, *J. Cardiovasc. Magn. Reson.* 22 (1) (2020) 19, doi:[10.1186/s12968-020-00610-6](https://doi.org/10.1186/s12968-020-00610-6).
- [16] A.S. Flett, et al., Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance, *JACC Cardiovasc. Imaging* 4 (2) (2011) 150–156, doi:[10.1016/j.jcmg.2010.11.015](https://doi.org/10.1016/j.jcmg.2010.11.015).
- [17] R. Karim, et al., Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement MR images, *Med. Image Anal.* 30 (2016) 95–107, doi:[10.1016/j.media.2016.01.004](https://doi.org/10.1016/j.media.2016.01.004).
- [18] Y. Wu, Z. Tang, B. Li, D. Firmin, G. Yang, Recent advances in fibrosis and scar segmentation from cardiac MRI: a state-of-the-art review and future perspectives, *Front. Physiol.* 12 (2021) 709230, doi:[10.3389/fphys.2021.709230](https://doi.org/10.3389/fphys.2021.709230).
- [19] O. Ronneberg, P. Fischer, T. Brox, U-Net: Convolutional Networks For Biomedical Image Segmentation, 9351, Springer, 2015 [Online]. Available., doi:[10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [20] G. Yang, et al., Multiview sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images, in: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 1123–1127, doi:[10.1109/EMBC.2018.8512550](https://doi.org/10.1109/EMBC.2018.8512550).
- [21] Z. Zhang, et al., Automatic segmentation of the left atrium from LGE-MRI based on U-Net and bidirectional convolutional LSTM, in: Proceedings of the Computing in Cardiology Conference, 2020, doi:[10.22489/CinC.2020.288](https://doi.org/10.22489/CinC.2020.288).
- [22] S. Moccia, et al., Automated scar segmentation from cardiac magnetic resonance-late gadolinium enhancement images using a deep-learning approach, in: Proceedings of the Computing in Cardiology Conference, 2018, doi:[10.22489/CinC.2018.278](https://doi.org/10.22489/CinC.2018.278).
- [23] A.S. Fahmy, et al., Three-dimensional deep convolutional neural networks for automated myocardial scar quantification in hypertrophic cardiomyopathy: a multicenter multivendor study, *Radiology* 294 (1) (2020) 52–60, doi:[10.1148/radiol.2019190737](https://doi.org/10.1148/radiol.2019190737).
- [24] F. Zabihollahy, J.A. White, E. Ukwatta, Fully automated segmentation of left ventricular myocardium from 3D late gadolinium enhancement magnetic resonance images using a U-Net convolutional neural network-based model, in: Proceedings of the Medical Imaging: Computer-Aided Diagnosis, San Diego, United States, 2019, p. 120, doi:[10.1117/12.2512146](https://doi.org/10.1117/12.2512146).
- [25] S. Moccia, et al., Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images, *Magn. Reson. Mater. Phys. Biol. Med.* 32 (2) (2019) 187–195, doi:[10.1007/s10334-018-0718-4](https://doi.org/10.1007/s10334-018-0718-4).
- [26] P. Kellman, et al., Dark blood late enhancement imaging, *J. Cardiovasc. Magn. Reson.* 18 (1) (2017) 77, doi:[10.1186/s12968-016-0297-3](https://doi.org/10.1186/s12968-016-0297-3).
- [27] *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer Berlin Heidelberg, New York, NY, 2015.
- [28] H.M.J.M. Nies, et al., Histopathological validation of semi-automated myocardial scar quantification techniques for dark-blood late gadolinium enhancement magnetic resonance imaging, *Eur. Heart J. Cardiovasc. Imaging* (2022), doi:[10.1093/ehjci/jeac107](https://doi.org/10.1093/ehjci/jeac107).
- [29] C.R. Harris, et al., Array programming with NumPy, *Nature* 585 (7825) (2020) 357–362, doi:[10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [30] A. Paszke, et al., PyTorch: an imperative style, high-performance deep learning library, in: Proceedings of the Advances in Neural Information Processing Systems, 32, 2019 [Online]. Available.:
- [31] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1026–1034, doi:[10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- [32] R. Karim, et al., Image data analysis for quantifying scar transmuralities in MRI phantoms for cardiac resynchronization therapy, in: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 1111–1114, doi:[10.1109/EMBC.2018.8512448](https://doi.org/10.1109/EMBC.2018.8512448).
- [33] R. Francis, et al., Prospective comparison of novel dark blood late gadolinium enhancement with conventional bright blood imaging for the detection of scar, *J. Cardiovasc. Magn. Reson.* 19 (1) (2017) 91, doi:[10.1186/s12968-017-0407-x](https://doi.org/10.1186/s12968-017-0407-x).
- [34] T.L. Braber, et al., Cardiac imaging to detect coronary artery disease in athletes aged 35 years and older. A scoping review, *Scand. J. Med. Sci. Sports* 28 (3) (2018) 1036–1047, doi:[10.1111/sms.12974](https://doi.org/10.1111/sms.12974).