# Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge

Alain Lalande, Zhihao Chen, Thibaut Pommier, Thomas Decourselle, Abdul Qayyum, Michel Salomon, Dominique Ginhac, Youssef Skandarani, Arnaud Boucher, Khawla Brahim, et al.

▶ **To cite this version:**

Alain Lalande, Zhihao Chen, Thibaut Pommier, Thomas Decourselle, Abdul Qayyum, et al.. Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge. Medical Image Analysis, 2022, 79, pp.102428. 10.1016/j.media.2022.102428 . hal-03682606

**HAL Id: hal-03682606**
**https://hal.science/hal-03682606v1**

Submitted on 8 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Medical Image Analysis 79 (2022) 102428

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge

Alain Lalande[a,b,1,*], Zhihao Chen[c,1], Thibaut Pommier[d], Thomas Decourselle[e], Abdul Qayyum[a], Michel Salomon[c], Dominique Ginhac[a], Youssef Skandarani[a], Arnaud Boucher[a], Khawla Brahim[a,f,g], Marleen de Bruijne[h,i,j], Robin Camarasa[h,i], Teresa M. Correia[k,l], Xue Feng[m], Kibrom B. Girum[a], Anja Hennemuth[n,o,p], Markus Huellebrand[n,o], Raabid Hussain[a], Matthias Ivantsits[n], Jun Ma[q], Craig Meyer[m], Rishabh Sharma[r,s], Jixi Shi[c], Nikolaos V. Tsekos[s], Marta Varela[t], Xiyue Wang[u], Sen Yang[v], Hannu Zhang[n], Yichi Zhang[w], Yuncheng Zhou[x], Xiahai Zhuang[x], Raphael Couturier[c], Fabrice Meriaudeau[a]

[a] *ImViA Laboratory, University of Burgundy, Dijon, France*
[b] *MRI Department, University Hospital of Dijon, Dijon, France*
[c] *Femto-ST Laboratory, University of Franche-Comté, Belfort, France*
[d] *Cardiology Department, University Hospital of Dijon, Dijon, France*
[e] *CASIS - CArdiac Simulation & Imaging Software SAS, Quetigny, France*
[f] *National Engineering School of Sousse, University of Sousse, Sousse, Tunisia*
[g] *LASEE laboratory, National Engineering School of Monastir, University of Monastir, Monastir, Tunisia*
[h] *Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, the Netherlands*
[i] *Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands*
[j] *Department of Computer Science, University of Copenhagen, Copenhagen, Denmark*
[k] *Centre of Marine Sciences, University of Algarve, Faro, Portugal*
[l] *School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom*
[m] *Department of Biomedical Engineering, University of Virginia, Charlottesville, USA*
[n] *Charité - Universitätsmedizin Berlin, Berlin, Germany*
[o] *Fraunhofer MEVIS, Bremen, Germany*
[p] *German Centre for Cardiovascular Research, Berlin, Germany*
[q] *Department of Mathematics, Nanjing University of Science and Technology, Nanjing, China*
[r] *Data Analysis and Intelligent Systems Lab, Department of Computer Science, University of Houston, Houston, USA*
[s] *Medical Robotics and Imaging Lab, Department of Computer Science, University of Houston, Houston, USA*
[t] *National Heart and Lung Institute, Imperial College London, London, United Kingdom*
[u] *College of Computer Science, Sichuan University, Chengdu, China*
[v] *College of Biomedical Engineering, Sichuan University, Chengdu, China*
[w] *School of Biological Science and Medical Engineering, Beihang University, Beijing, China*
[x] *School of Data Science, Fudan University, Shanghai, China*

ARTICLE INFO

*Article history:*
Received 1 July 2021
Revised 2 March 2022
Accepted 18 March 2022
Available online 24 March 2022

*Keywords:*
DE-MRI
Myocardium
Infarction
CNN

ABSTRACT

A key factor for assessing the state of the heart after myocardial infarction (MI) is to measure whether the myocardium segment is viable after reperfusion or revascularization therapy. Delayed enhancement-MRI or DE-MRI, which is performed 10 min after injection of the contrast agent, provides high contrast between viable and nonviable myocardium and is therefore a method of choice to evaluate the extent of MI. To automatically assess myocardial status, the results of the EMIDEC challenge that focused on this task are presented in this paper. The challenge's main objectives were twofold. First, to evaluate if deep learning methods can distinguish between non-infarct and pathological exams, i.e. exams with or without hyperenhanced area. Second, to automatically calculate the extent of myocardial infarction. The publicly available database consists of 150 exams divided into 50 cases without any hyperenhanced area after

* Corresponding author at: ImViA laboratory, Faculty of Medicine, University of Burgundy, 7 Bld Jeanne d'Arc, 21000 Dijon, France.
*E-mail address:* alain.lalande@u-bourgogne.fr (A. Lalande).
[1] Alain Lalande and Zhihao Chen contributed equally to this work.

https://doi.org/10.1016/j.media.2022.102428
1361-8415/© 2022 Elsevier B.V. All rights reserved.

injection of a contrast agent and 100 cases with myocardial infarction (and then with a hyperenhanced area on DE-MRI), whatever their inclusion in the cardiac emergency department. Along with MRI, clinical characteristics are also provided. The obtained results issued from several works show that the automatic classification of an exam is a reachable task (the best method providing an accuracy of 0.92), and the automatic segmentation of the myocardium is possible. However, the segmentation of the diseased area needs to be improved, mainly due to the small size of these areas and the lack of contrast with the surrounding structures.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

Myocardial infarction (MI) can be defined as myocardial cell death secondary to prolonged ischemia. One crucial parameter to estimate the prognosis after myocardial injury and then to evaluate the state of the heart after MI is the viability of the considered segment, i.e. if the segment recovers its functionality upon revascularization.

From cardiac MRI, the viability can be evaluated thanks to the assessment of left ventricular end-diastolic wall thickness, the evaluation of contractile reserve, and the extent and the transmural nature of the infarction evaluated from delayed-enhancement MRI (DE-MRI) (Schinkel et al., 2007; Kim et al., 1999). DE-MRI is a powerful predictor of myocardial viability after coronary artery surgery, suggesting an important role for this technique in clinical viability assessment (Selvanayagam et al., 2004).

A preliminary challenge organized in 2012 (Karim et al., 2016) has been already dedicated to the automatic processing of DE-MRI. This challenge showed promising results, but also indicated that some improvements should be done for potential use in clinical practice. The published dataset was rather small (including fifteen human and fifteen porcine pathological cases), did not target a specific disease and no clinical data were associated. Similar challenge was performed on DE-MRI for the study of fibrosis at the level of the left ventricle (Xiong et al., 2021).

As part of the Emidec challenge (automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI, http://emidec.com/) organized during the MICCAI conference in 2020, the objective of our paper is first to compare the latest methodological developments in image processing, in particular deep learning approaches such as convolutional neural network (CNN), to automatically segment the DE-MRI exams (including non-infarct exams and pathological cases with myocardial infarction and with or without persistent microvascular obstruction (PMO)), and secondly, thanks to the images and associated clinical data, to automatically classify the exams as non-infarct or pathological with hyperenhanced area. One of the main strengths of our database is the association of clinical data with DE-MRI, simulating the routine workflow in emergency departments.

## 2. Materials and methods

### 2.1. DE-MRI and clinical information

The EMIDEC dataset contains patients admitted in cardiac emergency department with symptoms of a heart attack. This dataset is composed of 150 patients, each of which with a MRI exam and the associated clinical characteristics. The exam is a conventional one acquired at the University Hospital of Dijon (France) and done to study the left ventricle in case of symptoms of heart attack and is compound of kinetic and DE-MR images. For the DE-MRI, the images are acquired roughly 10 min after the injection of a gadolinium-based contrast agent. A series of short axis slices cover the left ventricle from the base to the apex, allowing the accurate evaluation of the extent of myocardial infarction. Then the number of acquired slices vary according to the length of the left ventricle. The pixel spacing is between $1.25 \times 1.25$ mm$^2$ and $2 \times 2$ mm$^2$ (according to the patient), with a slice thickness of 8 mm and an image gap of 10 mm.

The shift among slices due to the patient's breath hold was corrected and all the slices for one exam had been aligned according to the gravity centre of the epicardium (indeed, each individual slice acquisition requires to the patient to hold his breath, then the localisation of the heart inside the image could be shifted). Along with the MRI, clinical and patient information were also recorded. According to the French law, ethical approval number was not needed as data were retrospectively collected and untraceable. In particular, concerning the images, using the NIfTi format, all the administrative information included in the header was discarded. Moreover, the clinical information is not specific enough to retrieve a patient. The patient features are characterised in Table 2 (Lalande et al., 2020). A patient was considered to be in overweight when the body mass index (BMI) is higher than 25. The history of coronary artery disease is positive when there is a previous acute cardiac event. The study of the electrocardiogram (ECG) allows classifying the heart attack as STEMI (ST-elevation myocardial infarction) type or not. STEMI-like myocardial infarction is the most serious type of heart attack, which is characterized by a long interruption of blood supply. A troponin test measures the levels of troponin T or troponin I proteins in the blood. These proteins are released during a myocardial infarction. Another biological marker is the NT-pro-brain natriuretic peptide (NTProBNP) measured in venous blood, and it is an indicator for the diagnosis of heart failure (Cochet et al., 2004). The left ventricular ejection fraction (LVEF) is calculated in the emergency department from echocardiography during the reception of the patient. Finally, Killip max corresponds to the maximum Killip score, which is a classification based on the physical examination of patients with possible acute MI (Killip and Kimball, 1967). We can remark that the group of patients with non-pathological exams provides also abnormal values, because the mean LVEF is lower than 50%, the troponin and NT-ProBNP levels are higher than normal. Concerning the NTProBNP, we can notice that the mean value is higher than for the patholog-

**Table 1**
List of the challengers.

| Challenger teams | Segmentation contest | Classification contest |
|---|---|---|
| Brahim et al. (2021) | x | |
| Camarasa et al. (2021) | x | |
| Feng et al. (2021) | x | |
| Girum et al. (2021) | x | x |
| Huellebrand et al. (2021) | x | |
| Ivantsits et al. (2021) | | x |
| Lourenço et al. (2021) | | x |
| Sharma et al. (2021) | | x |
| Shi et al. (2021) | | x |
| Yang and Wang (2021) | x | |
| Zhang (2021) | x | |
| Zhou et al. (2021) | x | |

**Table 2**
Characteristics of the patient features for non-infarct exams and infarct exams.

| Patient feature | non-infarct exams ($n = 50$) | infarct exams ($n = 100$) |
|---|---|---|
| Sex* | 38 females and 12 males | 23 females and 77 males |
| Age* (years) [min-max] | $66 \pm 14$ [27–88] | $59 \pm 12$ [29–89] |
| Tobacco (yes, no, former smoker)* | 18%, 22%, 60% | 44%, 21%, 35% |
| Overweight[a] | 62% | 53% |
| Arterial hypertension* | 58% | 31% |
| Diabetes | 20% | 10% |
| History of coronary artery disease | 4% | 12% |
| ECG (ST elevation)* | 30% | 80% |
| Troponin (ng per mL)* [min-max] | $7.68 \pm 12.91$ [0–83] | $101.04 \pm 101.35$ [0.3–420] |
| Killip max (1,2,3,4) | 76%, 22%, 2%, 0% | 83%, 12%, 2%, 3% |
| LVEF[b] (percentage) [min-max] | $49.62 \pm 13.49\%$ [25–70] | $47.74 \pm 13.17\%$ [20–70] |
| NTProBNP[c] (pg per mL) [min-max] | $2136 \pm 3696$ [3–22577] | $1314 \pm 2109$ [14–9970] |

[a] If BMI $> 25$.
[b] Left Ventricular Ejection Fraction, calculated from transthoracic echocardiography.
[c] N-terminal pro-B-type natriuretic peptide.* $p < 0.05$ between the two groups from an independent $t$-test (null hypothesis rejected).

**Table 3**
DE-MRI evaluations of non-infarct and infarct exams according to the manual annotations. This table lists the characteristics of different tissues in the DE-MRI for the whole dataset.

| Tissue | Non-infarct exams ($n = 50$) | | | Infarct exams ($n = 100$) | | |
|---|---|---|---|---|---|---|
| | Volume (mL)[a] | PIM (%)[b] | Presence (%)[c] | Volume (mL)[a] | PIM (%)[b] | Presence (%)[c] |
| Myocardium | 96.32 ± 22.07 | – | – | 119.28 ± 32.28 | – | – |
| Left ventricular cavity | 83.32 ± 25.27 | – | – | 128.87 ± 48.17 | – | – |
| Myocardial infarction[d] | 0 | 0 | 0 | 23.55 ± 19.28 | 18.25 ± 11.52 | 100 (79.78) |
| PMO | 0 | 0 | 0 | 2.34 ± 5.14 | 1.65 ± 3.03 | 51 (23.27) |

[a] Absolute tissue volume per case.
[b] Percentage of Infarcted Myocardium. This index is reserved for myocardial infarction and PMO.
[c] Percentage of cases where tissue is present, while the value in brackets gives the percentage of slices.
[d] The PMO is included.

ical exams, and the standard deviation is very high, confirming the variability of these cases.

### 2.2. Dataset and contests

The overall dataset consists of 150 exams, with 100 cases for the training (67 pathological cases (or infarct cases) and 33 non-infarct cases, where ground truths are provided) and 50 cases for the testing (33 infarct cases, 17 non-infarct cases). A case is defined as non-infarct if there is no hyperenhanced signal in any slice on the DE-MRI data, knowing that the patient could suffer from another cardiovascular disease than MI and then get abnormal clinical parameters. The dataset does not include any healthy volunteers for the non-infarct cases, as all patients were having cardiac issues and data were collected at the emergency department of the hospital. Therefore, a case is defined as infarct exam if there is an hyperenhanced signal on at least one slice. Alone this abnormal signal cannot discriminate between fibrosis and MI and clinical information is required to make this distinction. Each exam is divided into two parts, a DE-MRI exam composed of a series of short-axis slices and the associated clinical information (Lalande et al., 2020). For each image, the contours of the myocardium, as well as the contours of the infarcted area and the PMO areas, if present, are considered as the ground truths, allowing the calculation of the main clinical metrics considering the whole slices for one exam. This contouring was performed firstly by an experienced user (a cardiologist with 10 years of experience in cardiology and MRI) with the QIR software (CASIS, Quetigny, France) and then corrected if necessary by a second expert (a well-trained biophysicist with 20 years of experience). During the contouring of the endocardial border, the papillary muscles were included in the cavity. The experts were requested to draw black areas surrounded by a bright area as PMO areas while discarding black signal due to noise or artefacts. Tissue characteristics according to the manual annotations can be found in Table 3. Along with MRI, the clinical and physiological characteristics are provided. The EMIDEC challenge contains two independent contests: the segmentation challenge and the classification challenge. The goal of the segmentation contest is to compare the performance of automatic methods on the segmentation of the myocardium for all the DE-MRI exams, as well as the segmentation of the myocardial infarction and PMO areas on exams classified as infarct ones. The goal of the classification contest is to classify the exams as non-infarct or infarct, according to the clinical data with or without the DE-MRI exams (two sub-challenges). In order to avoid any bias between the two contests, the order of the cases is different in the testing set, and moreover, new cases were randomly added (and similarly some were removed) for the classification contest. Table 1 summarizes the different participants of the EMIDEC challenge.

### 2.3. Evaluation metrics

For the segmentation contest, the clinical metrics are the most widely used in cardiac clinical practice, i.e. the average errors for the volume of the myocardium of the left ventricle (mL), the volume (mL) and the percentages of MI and PMO. The geometrical metrics are the average Dice index for the different areas and the Hausdorff distance (in 3D) for the myocardium. For each metric, a ranking is done, and the final ranking consists of the sum of the ranking for each metric. To better evaluate the segmentation results of the PMO, the case-wise and slice-wise accuracies are additionally calculated, but were not taken into account for the challenge ranking. For the classification contest, only classification accuracy was used.

## 3. Evaluated architectures

### 3.1. Segmentation contest

The main objective of the segmentation contest is to automatically provide the contours of the myocardium on each slice, as well as the delineation of the diseased areas.

#### 3.1.1. Image preprocessing and data augmentation

To ensure that semantic information in DE-MRI can be efficiently interpreted by the segmentation models, some challengers employed adaptive image preprocessings on the raw MR images. For example, image normalization aims at correcting the heterogeneous intensity between cases. Yang and Wang and Feng et al. applied the *Z*-score normalization on each slice with the following formula:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where $z$ is the pixel intensity after the *Z*-score normalization, $\mu$ the mean intensity at the level of the MR slice and $\sigma$ the standard deviation of the slice intensity. Normalized images have a grey level distribution with zero mean and unit standard deviation so that the inter-case intensity distribution is uniform.

Exams have some slight inconsistencies in the plane dimensions. In order to ensure a uniform input shape of the predictive models, challengers processed the plane dimensions of the input data differently. The first type of method is cropping, e.g. Feng et al. cropped a fixed size in the centre of each slice. In addition to the cropping, a linear interpolation was also performed to resize the images to a uniform shape (Camarasa et al.). Besides the processing on the slice shape, some challengers also interpolated the image to have a consistent pixel spacing. Thanks to the alignment of the slices according to the gravity center of the epicardium, no additional preprocessing concerning the relative inter-slice position is needed if a 3D predictive model is employed by the challengers.

The amount of training data directly affects the performance of supervised models. A reasonable data augmentation method can equivalently expand the size of the training set. Camarasa et al. performed rotations, elastic deformations, and flips on slices to randomly produce supplementary training data while Feng et al. forced the model to ignore the specificity for different orientation features by the rotations only. Lourenço et al. adjusted the original semantic information by adding stochastic noise, applying k-space corruption, small image rotations, intensity scalings, and smooth non-rigid deformations. Zhou et al. proposed another data augmentation method that was based on the mix-up strategy (Zhang et al., 2018). The mix-up strategy constructs virtual training examples as follows:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{2}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{3}$$

where $x_i$ and $x_j$ are raw input vectors, $y_i$ and $y_j$ are one-hot label encodings, $\tilde{x}$ and $\tilde{y}$ is the pair of artificially created data. $\lambda$ is a coefficient belonging to [0,1]. Based on this approach, Zhou et al. made a targeted improvement to make the generated images closer to a blend of two adjacent images. The proposed mix-up formula for the MRI augmentation is:

$$\tilde{x} = \lambda x_i + (1 - \lambda) T x_j \tag{4}$$

where $T$ denotes an affine transformation, and accordingly the similar formula for the mask data augmentation. Given the greater focus on the ROI (Region Of Interest corresponding to the myocardium), the affine transformation $T$ tries to fit the transformation from the foreground area (LV+Myocardium) in a randomly chosen slice $x_i$ to the foreground area in another randomly chosen slice $x_j$. In the affine transformation, the scaling factor, i.e. the linear map is $[s, s]^\top$ where $s = l^i/l^j$, $l^i$ and $l^j$ are the average distance from the foreground pixels to the foreground center for the slice $i$ and the slice $j$, respectively. The translation offset is $[c_x^i - c_x^j, c_y^i - c_y^j]^\top$ where $c_x$ and $c_y$ denote the coordinates of the foreground area centre. Thus, the matrix of $T$ is:

$$\begin{pmatrix} s & 0 & c_x^i - s \cdot c_x^j \\ 0 & s & c_y^i - s \cdot c_y^j \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

#### 3.1.2. Segmentation frameworks

Challengers employed segmentation frameworks with a different number of stages. Most of the challengers first delineated the myocardium (endocardial and epicardial borders) and then segmented the different myocardial tissues in the ROI corresponding to the myocardium with another model in a second step. Other challengers proposed one-stage models to obtain an end-to-end segmentation of all the target tissues. Zhang proposed the cascaded 2D–3D framework where the 2D-model's receptive field was limited to intra-slice for preliminary segmentation, then the cascaded 3D-model took the 2D preliminary segmentation mask and the whole volume for the fine segmentation. This conception aims at restricting the impact of intra-slice heterogeneity and taking into account the volumetric information for the more accurate segmentation. The network configurations are inspired by nnUnet (Isensee et al., 2021). Fig. 1 shows the architecture of the cascaded two-stage framework. Camarasa et al. also employed a usual two-stage segmentation pipeline but the scar segmentation was uncertainty-based: the ROI segmented by the first model passed through a probabilistic Auto-Encoder using Monte–Carlo dropout. The generated uncertainty map corresponding to the segmented ROI by the Auto-Encoder was then fed into the second model for the scar segmentation. This proposal was intended to assess whether the method could increase the attention on rare examples that are otherwise poorly segmented.

#### 3.1.3. U-Net-based encoding-decoding models

The semantic image segmentation task can be usually treated with encoding-decoding models. Most challengers employed U-Net-based models (Ronneberger et al., 2015) motivated by its success in many medical image segmentation work. In this subsection, the details of all the employed U-Net-based models will be introduced.

*Building blocks.* The vanilla U-Net employs the conventional convolution-pooling architecture as the basic encoding block. To better interpret the semantic information, challengers attempted with more recent blocks of CNNs. In the encoding branch, Yang and Wang and Girum et al. applied the Squeeze-and-Excitation (SE) block (Hu et al., 2018) to better model the interdependencies between channels of the convolutional features.

On the decoding side, Inverted Residual Blocks (IRB) were employed by Brahim et al. Selective Kernel (SK) (Li et al., 2019) was another block employed in the decoding side by Yang and Wang.

In addition to the above-featured building blocks, challengers also tried other relatively more common blocks such as the residual block of the ResNet (He et al., 2016) and its aggregated variant ResNeXt (Xie et al., 2017), as well as the Inception module where convolutions of different receptive field interpret input features at the same time (Szegedy et al., 2017). The attention block (Oktay et al., 2018) was also mentioned by several challengers to focus on valuable features at the skip connections and the bottleneck. The illustration of the featured building blocks can be found in Fig. 2 and more information is given about them in the supplementary material.

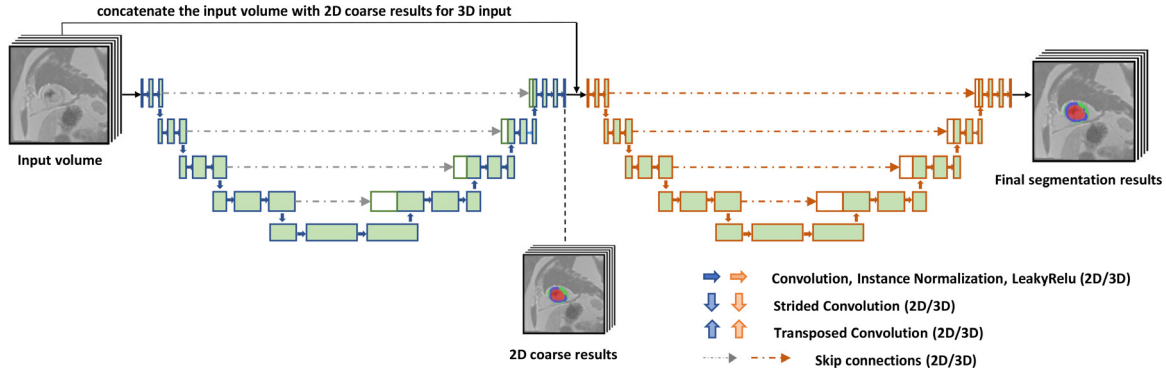


**Fig. 1.** Cascaded 2D-3D framework of Zhang for the myocardial tissue segmentation. The 2D network on the left performs a preliminary segmentation focusing on intra-slice information. The 3D network on the right then takes the MRI and the obtained segmentation mask as its input.

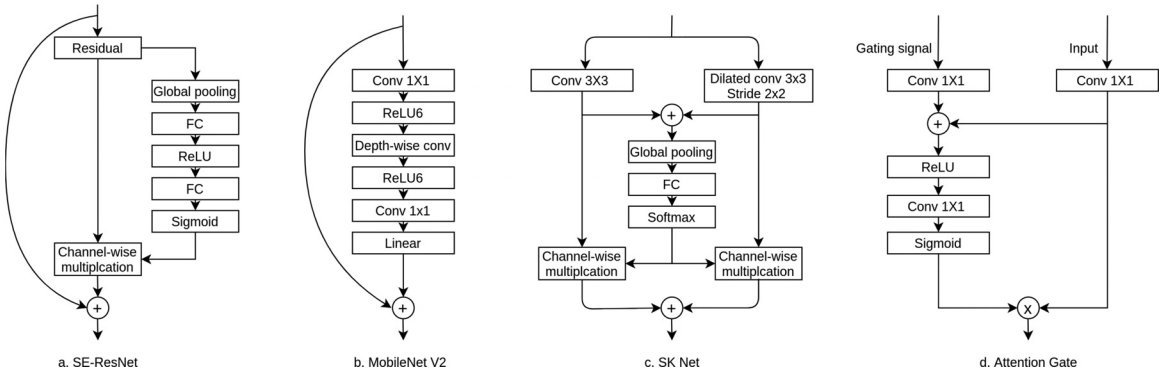


**Fig. 2.** Featured building blocks for the segmentation networks. The conceptions come from the original papers of the challengers or are slightly modified to better adapt the contest. a. SE-ResNet: the residual SE block, b. MobileNet V2: the IRB from MobileNet V2, c. SK Net: the SK module can be deployed in the encoding or decoding phases, d. Attention Gate: the attention gating from Attention U-Net should be deployed at the skip connection. The Gating signal comes from the encoding side and the input signal denotes the up-sampled features from the decoding side. The first two conv $1 \times 1$ ensure the same number of channels for the two signals of the Attention Gate.

Challengers also reported the use of a variety of activation functions. Like in most of the current deep learning models, the activation functions themselves are all non-linear equations, their core functionality is to ensure that nontrivial problems can be fitted by deep neural networks. Therefore, sigmoid, rectifier (ReLU) (Glorot et al., 2011) and its leaky variant, exponential linear (ELU) (Clevert et al., 2016), Swish (Ramachandran et al., 2017), etc. activation functions were employed. Fig. 3 illustrates the deployed activation functions by challengers.

*Loss functions and penalizations.* The category imbalance is significant in the challenge dataset, that is, the myocardial infarction and the PMO have few instances in terms of the number of pixels as shown in Table 3. To address this issue, challengers investigated different loss functions such as cross entropy loss, weighted cross entropy loss and Dice loss (see supplementary material for more details). It can be observed that for the scar segmentation, the categorical cross entropy loss was weighted (Zhang, Yang and Wang, Zhou et al.) while the multi-class Dice loss was not weighted (Camarasa et al., Yang and Wang, Zhang). To leverage the cross entropy loss and Dice loss, their combination termed Comboloss (Taghanaki et al., 2019) was also practiced by many challengers (Girum et al., Yang and Wang, Zhang) for the ROI or the whole tissues segmentation.

Apart from the loss functions that penalize the difference between the target and the prediction, other prior information-based penalizations were investigated by challengers. Brahim et al. applied the 3D auto-encoder as a part of the loss to refine the mask contours. The employment of the auto-encoder with the cardiac MRI was first proposed by Yue et al. (2019) for the myocardium segmentation. In the original work, the auto-encoder learns the 2D shape prior of the myocardium since the short-axis view of the left ventricle should be a closed circle except for the extreme apical and basal slices. The auto-encoder can be thought as an annex network following the segmentation network so that the loss of the auto-encoder takes part of the backpropagation. Similarly, with reference to the prior anatomical knowledge, Zhou et al. proposed the neighborhood penalty as a weak constraint strategy. Given the fact that the PMO should be in contact with the infarction and the whole scar area inside the myocardium, this penalty encourages such correlated tissues to stick together.

*Inter-slice and intra-slice information.* The cardiac MR images can be considered as pseudo 3D data, i.e. the voxel spacing is inconsistent between the in-plane and between planes. Challengers adopted different strategies to deal with the inter- and intra-slice information. The first one omits the inter-slice correlation, i.e. all the tissues are segmented from single slices whether the framework is one-stage or two-stage (Feng et al., Girum et al., Huellebrand et al., Zhou et al.). The second one only takes 3D inputs while the data format organization is different. Camarasa et al. employed a 3D CNN where the convolution kernel was 3D. Yang and Wang treated multi-slice data as different channels, i.e. at the input layer each channel stocked one slice and the following convolutions were 2D. The major difference of the 3D data interpretation between the 3D convolution and the 2D convolution with RGB channels-like inputs is the relative positional information between the slices. The 2D convolution cannot distinct the slice order while the 3D convolution retains the inter-slice information as local vector data. The last strategy is a compromise approach (Brahim et al., Zhang): the ROI or preliminary segmentation only refers to the intra-slice information and the scar or final segmentation consid-

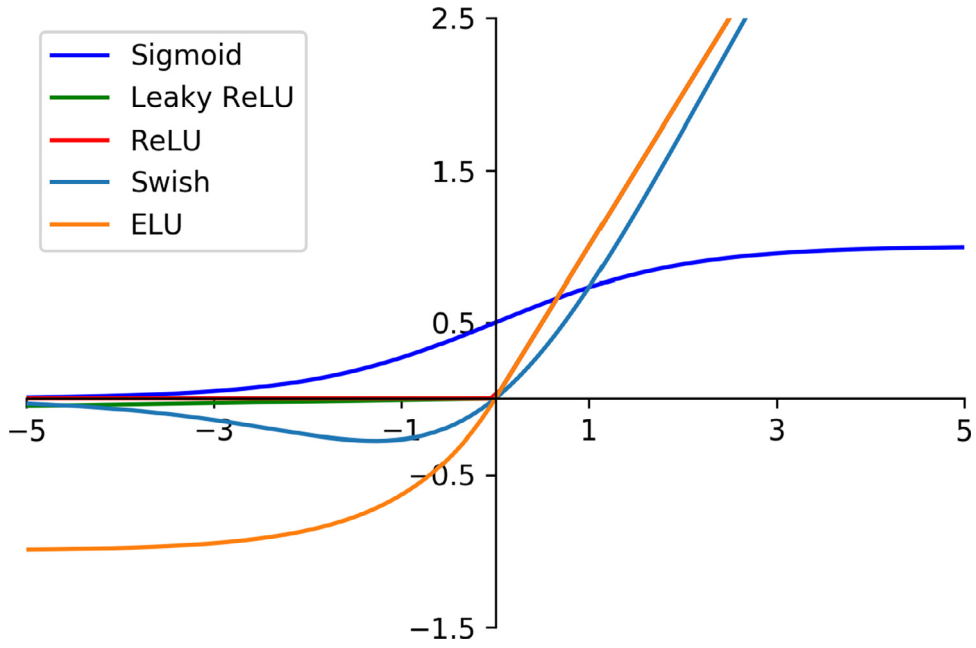


**Fig. 3.** Deployed non-linear activation functions in the neural networks for the segmentation task.

ers both the intra- and inter-slice information. The purpose is to avoid the potential inter-slice heterogeneity for the myocardium or preliminary segmentation, and take the advantage of the inter-slice information for the scar since the recognition of different myocardial tissues relies more on their neighbouring slices.

*3.1.4. Mixture model for the scar segmentation*

Apart from the U-Net-based models that most challengers employed, a mixture model was proposed by Huellebrand et al. for the scar segmentation. The application of the mixture model on the cardiac MRI was inspired by the work of Hennemuth et al. (2013). The mixture model differs the scar tissues only according to the intensity distribution. The challengers trialed the mixture of a Rician and a Gaussian distribution and the mixture of Rayleigh and Gaussian distribution, and then adopted the latter which was proved better fitted to the scar tissues in the DE-MRI. Finally, inspired by Hennemuth et al. (2008), a watershed segmentation in high-intensity voxels was used at the inner part of the myocardium to get the segmented contours.

*3.1.5. Post-processing*

According to prior information, challengers employed simple post-processing methods. Huellebrand et al. proposed a thresholding for the segmented PMO: assuming that the PMO should be in contact with the cavity or the infarction, the contours detected by morphological closing were removed from the raw segmentation of the PMO. Zhang adopted another simple treatment that removed all the scattered pixels from the segmentation.

*3.2. Classification contest*

The objective of the classification contest is to classify each exam as non-infarct or infarct, whatever the extent of the myocardial infarction.

*3.2.1. Basic data interpretation algorithms*

Challengers employed a variety of machine learning-based algorithms to interpret the DE-MRI and the clinical features. Provided with the MRI, a simple down-sampling CNN as AlexNet (Krizhevsky et al., 2012) encodes the images to regression or classification outputs (Lourenço et al., Sharma et al., Shi et al.), or optionally U-Net based down-sampling up-sampling models yield the segmentation of different myocardial tissues so that the volume of each tissue can be quantified (Girum et al., Lourenço et al.).

To interpret the textual data of the clinical and physiological information, the choice of predictive models is more diverse. The common functionality is its ability to solve non-linearly separable problems. For example, the MultiLayer Perceptron (MLP) (Hinton et al., 2006) is a feedforward artificial neural network. Inputs are passed through multiple layers in which data are mapped with non-linear activation functions in the forward stage (Ivantsits et al., Sharla et al.). The decision tree (Quinlan, 1987) and the random forest (Ho, 1995) are flow-chart-like decision models that consist of nodes (Ivantsits et al., Sharma et al., Shi et al.). The random forest corrects the overfitting habit of the decision trees by training uncorrelated trees and the final decision is made by individual trees. Boosting methods are the ensemble of sequentially connected weak learners (Breiman, 1996). In the context of decision trees, the gradient boosting decision trees build a series of trees, which are the weak learners in this boosting method. Errors are passed between trees, with each tree attempting to reduce the errors passed from the previous tree (Friedman, 2001) (Ivantsits et al.). Moreover, usual statistical models such as Support Vector Machine with non-linear kernel (Scholkopf, 2001), k-Nearest Neighbors (Fix and Hodges, 1989), the logistic regression (Peng et al., 2002) were investigated by challengers (Girum et al., Ivantsits et al., Sharma et al.).

*3.2.2. Data fusion and decision about the presence of myocardial infarction*

The classification contest allows challengers to take advantages of both the DE-MRI and the clinical and physiological data to make the automatic decision. However, the different format and dimension between the images and the textual data constrain the decision with a single predictive model. Data fusion is therefore a challenging issue to achieve the maximum semantic information. Girum et al., Lourenço et al. and Shi et al. deployed the same strategy of predicting the volumes of different tissues as additional textual features alongside the 12 clinical and physiological fea-

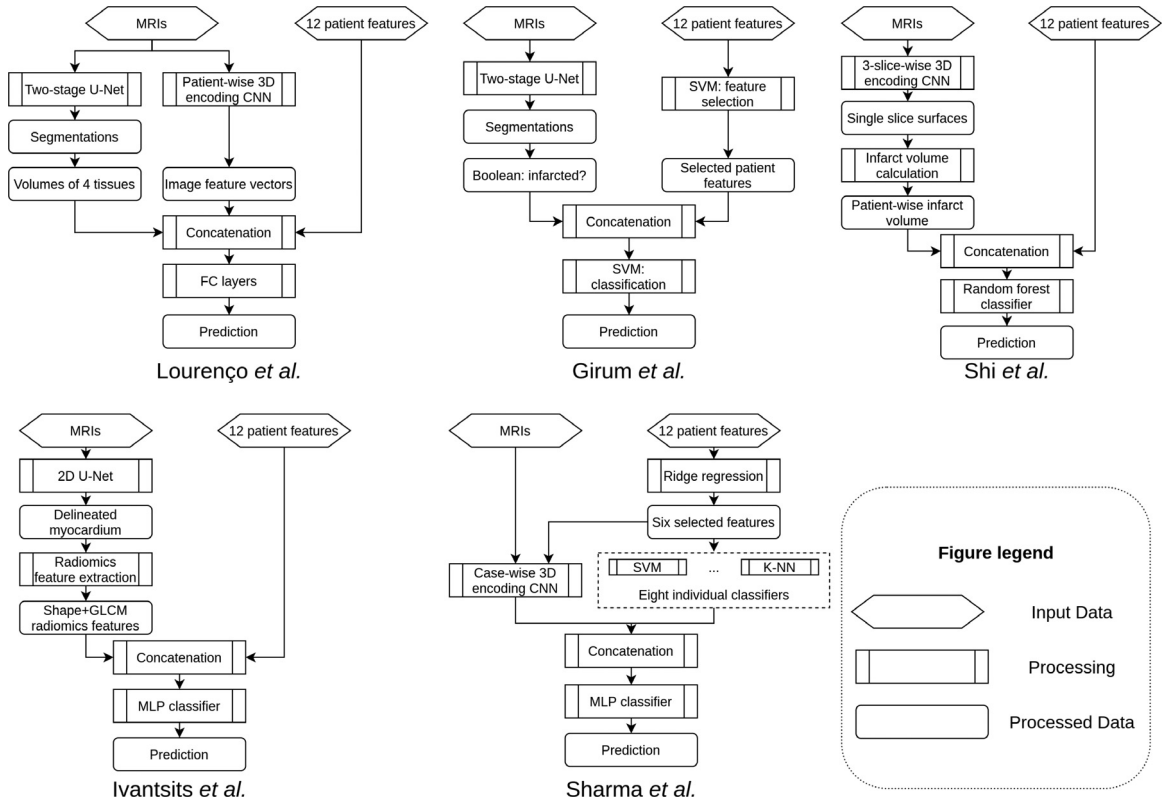


**Fig. 4.** Proposed multi-input classification pipelines by challengers for the classification contest.

tures. Nevertheless, the volume estimation and the decision making models are different among these approaches. Lourenço et al. and Girum et al. employed U-Net-based models to get the segmentation, while Shi et al. performed an encoding CNN to directly get the surface regression. Apart from the surface regression methods, the concatenation of the surface information to other textual features was also variable. Lourenço et al. added the volumes of all myocardial tissues as four additional textual features. Girum et al. only considered if the case is an infarct exam as one additional Boolean feature and Shi et al. referred to the infarction volume as one additional numerical feature. Ivantsits et al. tried to interpret the DE-MRI as textual information as well. However, the obtained textual information was radiomic features (Cetin et al., 2018) instead of the volume of the tissues. The radiomic features interpreted from the DE-MRI were intended to model the myocardial features such as the intensity, shape, and spatial characteristics. In practice, Ivantsits et al. investigated the shape and the Gray Level Co-occurrence Matrix (GLCM) that described the second-order joint probability function of an image region as the experimental radiomic features. Sharma et al. proposed a stacked multi-modal approach without obtaining intermediate data such as the infarct volume or the radiomics features. The classifications were first achieved by a series of statistical models and a multi-modal CNN. Then the individual classifications were fed into an MLP to get the final decision. The application of the series of classification models could be thought as a boosting method and the models inside played the role of weak classifiers since their decisions would be judged together with the CNN's output by the MLP at the end of the proposal. All the diagrams of the classification pipelines can be found in Fig. 4.

**Table 4**
Principal concepts of the methods for the segmentation contest.

| Challenger(s) | Framework | Methods | Highlights |
|---|---|---|---|
| Brahim et al. | Two-stages | Myocardium: 2D U-Net with Attention and IRB<br>Infarct: 3D U-Net variant | 3D Auto-encoder to perfect myocardium shape |
| Camarasa et al. | Two-stages | Myocardium: 3D U-Net variant<br>Infarct: 3D U-Net variant | Uncertainty myocardial area generated by probabilistic auto-encoder for infarct segmentation |
| Feng et al. | One-stage | 2D U-Net with dilated convolution | Data augmentation with additional scar tissues |
| Girum et al. | Two-stages | Myocardium: 2D U-Net with SE block<br>Infarct: 2D U-Net with SE block | Independent myocardium and infarct segmentation from non-cropped MRI |
| Huellebrand et al. | Two-stages | Myocardium: 2D U-Net variant<br>Infarct: mixture model | Transfer learning with cine-MRI<br>Post-processing with thresholding and morphological closing |
| Yang et al. | One-stage | 2D U-Net with SE and SK blocks | RGB channel-like adjacent slices input<br>Two decoder branches supervised by myocardium and infarct masks |
| Zhang | Two-stages | Preliminary: 2D U-Net variant<br>Definitive: 3D U-Net variant | 3D MRI with cascaded 2D segmentation as 3D input |
| Zhou et al. | One-stage | 2D U-Net with Attention | Data augmentation with mix-up strategy Neighborhood penalty as neighboring loss |

**Table 5**
Results of the segmentation contest. The metrics are given by target tissue (myocardium, infarct and PMO). The table is sorted by the general ranking of the contest, which is calculated from the nine subranks. Best results in bold.

| Challenger(s) | Myocardium | | | Infarction | | | PMO | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dice | Vol. Diff. (mL) | Hausdorff (mm) | Dice | Vol. Diff. (mL) | Pct. Diff. (%)[a] | Dice | Vol. Diff. (mL) | Pct. Diff. (%)[a] | Acc. (case,%)[b] | Acc. (slice,%)[b] |
| Zhang | **0.879** ± 0.027 | **9.26** ± 9.08 | **13.01** ± 8.81 | **0.712** ± 0.268 | **3.12** ± 5.15 | **2.38** ± 0.031 | **0.785** ± 0.393 | **0.63** ± 2.27 | **0.38** ± 0.012 | **84.00** | **94.97** |
| Feng et al. | 0.836 ± 0.124 | 15.19 ± 16.41 | 33.77 ± 111.63 | 0.547 ± 0.340 | 3.97 ± 8.36 | 2.89 ± 0.045 | 0.722 ± 0.432 | 0.88 ± 3.41 | 0.53 ± 0.017 | 80.00 | 90.78 |
| Yang et al. | 0.855 ± 0.027 | 16.54 ± 10.27 | 13.23 ± 6.80 | 0.628 ± 0.315 | 5.34 ± 7.88 | 4.37 ± 0.062 | 0.610 ± 0.463 | 1.85 ± 3.32 | 1.69 ± 0.033 | 76.00 | 81.56 |
| Huellebrand et al. | 0.841 ± 0.051 | 10.87 ± 8.53 | 18.3 ± 15.74 | 0.379 ± 0.296 | 6.17 ± 8.36 | 4.93 ± 0.059 | 0.523 ± 0.483 | 0.95 ± 3.00 | 0.64 ± 0.015 | 70.00 | 85.75 |
| Camarasa et al. | 0.757 ± 0.111 | 17.11 ± 15.45 | 25.44 ± 21.71 | 0.308 ± 0.280 | 4.87 ± 8.49 | 3.64 ± 0.047 | 0.605 ± 0.485 | 0.87 ± 3.27 | 0.52 ± 0.016 | 74.00 | 84.36 |
| Zhou et al. | 0.825 ± 0.057 | 13.29 ± 11.34 | 83.42 ± 158.97 | 0.378 ± 0.309 | 6.10 ± 9.45 | 4.71 ± 0.06 | 0.520 ± 0.487 | 0.88 ± 3.38 | 0.54 ± 0.017 | 64.00 | 86.87 |
| Brahim et al.[c] | 0.791 ± 0.050 | 12.68 ± 10.59 | 23.87 ± 11.52 | 0.274 ± 0.379 | 7.05 ± 12.73 | 5.19 ± 0.074 | 0.641 ± 0.479 | 0.83 ± 3.109 | 0.50 ± 0.016 | 74.00 | 89.39 |
| Girum et al.[c] | 0.803 ± 0.057 | 11.81 ± 14.09 | 51.48 ± 98.15 | 0.340 ± 0.474 | 11.52 ± 16.53 | 8.58 ± 0.101 | 0.780 ± 0.414 | 0.89 ± 3.61 | 0.51 ± 0.018 | 78.00 | 89.66 |

[a] Pct. Diff. : Difference between the percentage of the infarcted myocardium.
[b] Additional metrics. These metrics were not taken into account in the ranking.
[c] Co-authors come from the challenge organization team. Do not participate in rankings.

**Table 6**
Segmentation of the MI considering the different coronary arteries. The difference of the PIM between the prediction and the ground truth is evaluated according to LAD, RCA and LCX. Best results in bold. There is no statistical difference between the arteries whatever the challenger(s) (ANOVA test).

| Challenger(s) | Difference in the PIM (%) | | |
|---|---|---|---|
| | LAD | RCA | LCX |
| Zhang | **3.31** | **1.90** | **2.02** |
| Feng et al. | 3.59 | 3.34 | 2.74 |
| Yang et al. | 5.56 | 3.62 | 2.64 |
| Huellebrand et al. | 4.37 | 4.69 | 5.14 |
| Camarasa et al. | 5.81 | 3.15 | 4.60 |
| Zhou et al. | 4.28 | 5.26 | 3.12 |
| Brahim et al. | 8.86 | 5.83 | 5.82 |
| Girum et al. | 12.00 | 7.51 | 5.93 |

## 4. Results

The results were obtained on the datasets used during the testing phase of the final challenge session.

### 4.1. Segmentation contest

Table 4 shows the key conception details of the segmentation contest challengers. In Table 5 the evaluation results of each target tissue are provided. Results reveal that the myocardium segmentation is globally satisfying while the infarction is relatively challenging to be correctly predicted. The metrics of Dice and volumes used during the challenge for the PMO segmentation may not be consistent since the PMO only represents a very small volume of the data. Indeed, a total absence of PMO on all the images seemingly provides correct results with Dice index or volumes. In contrast, the accuracy highlights the efficiency of the different methods to detect PMO areas. Moreover, segmentation results by slice position can be found in the supplementary material for all the participants. Apart from the evaluation by tissue for the whole myocardium, the impact of myocardium infarction by coronary artery are also assessed to better take into account the cardiac anatomy. Table 6 shows the difference of the percentage of infarcted myocardium between the predictions and the ground truth, according to the myocardial areas supplied by the Left Anterior Descending artery (LAD), the Right Coronary Artery (RCA) and the Left CircumfleX artery (LCX), considering the initial 17 segments' model of Cerqueira et al. (2002). One can notice that the worst results are generally obtained for the LAD and there is not a large difference between the results obtained for the RCA and LCX. To verify whether there is a statistical difference between the groups, a one-way ANOVA test was used and a $p$-value less than 0.05 was considered significant. The differences between the 3 groups are not significant, although there is a trend for Yang and Wang ($p = 0.06$) and Girum et al. ($p = 0.08$).

To intuitively present the state-of-the-art segmentation results and the challenges to be overcome, segmentation masks from different challengers on five typical slices are selected. Fig. 5 covers the selected MRI slices and theirs ground truth masks, showing for each slice two well-performed segmentations and two segmentations to optimize. Here are the details:

1. Slice A is close to the apex. Therefore only a small part of the right ventricle appears in this slice (blue arrow). Methods on rows 1 and 2 successfully delineated the junction between the left and the right ventricles, while method on row 3 overestimated the right ventricle and method on row 4 wrongly segmented the right ventricle as a small infarct (yellow arrows).

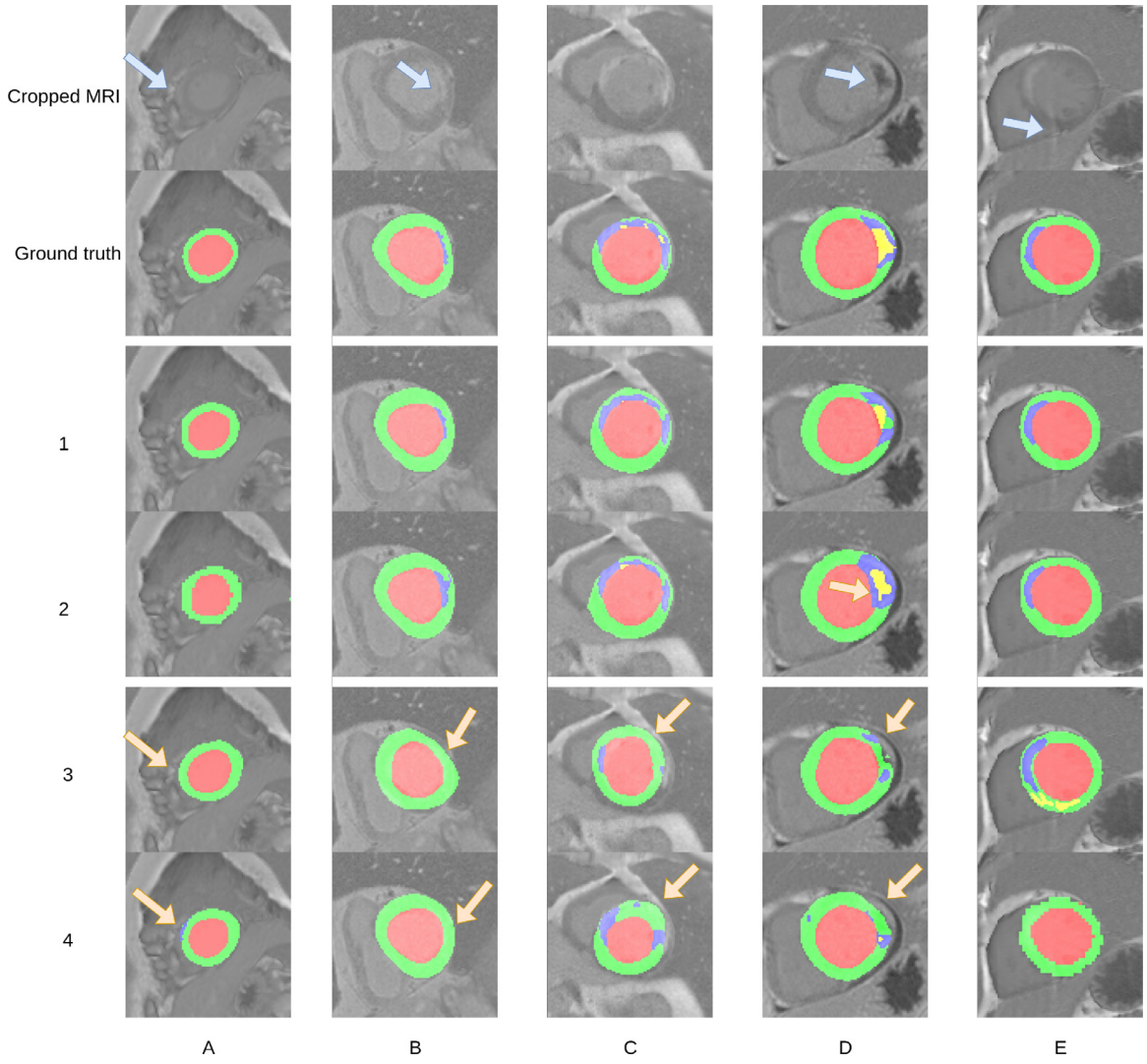


**Fig. 5.** Segmentation results on five challenging slices. Rows 1 and 2 denote satisfying segmentation results, rows 3 and 4 denote segmentations to be optimized. Columns A-E denote five slices from different testing cases. The masks in each row may come from different challengers. Blue arrows highlight difficult areas to detect (low contrast, presence of artifact, etc.). Yellow arrows show differences between challengers for specific segments. Cardiac cavity in red, normal myocardium in green, myocardial infarction in blue and PMO in yellow. See details in the text.

2. Slice B involves an infarct that connects the cavity (blue arrow). Methods on rows 3 and 4 failed due to the low contrast and narrow width of the infarct.
3. Scar tissues in slice C have a broken shape: On the upper side, the scar tissues and the normal myocardium intersperses. The interspersed area was wrongly segmented as normal myocardium on rows 3 and 4 (yellow arrows).
4. Slice D involves an important PMO area. Although the best adaptive approaches recognized the existence of the PMO, a part of the PMO area was segmented as the normal myocardium (row 1) or the infarct was over-estimated (row 2). Most of the other challengers wrongly segmented the infarct wrapping the PMO as the adipose tissue on the lateral segment of the myocardium (yellow arrows in rows 3 and 4).
5. Slice E involves an artifact (blue arrow). Reassuringly, for most challengers, the presence of this artifact on the myocardium did not interfere the segmentation while some challengers made atypical errors on this slice.

In addition, the segmentations of all challengers on all the slices of one entire exam are provided in the supplementary material.

### 4.2. Classification contest

The classification contest results are listed in Table 7. The best results were achieved on the merged textual and graphical data. Lourenço et al., Girum et al. and Shi et al. also submitted their classification results relying on sole textual data. The achieved accuracy on the textual data were 70%, 78%, and 74% respectively, which were significantly outperformed by their model with data fusion in Table 7 (82%, 82%, and 92% respectively). The best method failed only on 3 cases among 50, which we can consider as an excellent result.

## 5. Discussion

### 5.1. Challenge results

The overall challenge results were satisfactory. For the segmentation task, the best method obtained a Dice score of 0.879 for the myocardium and of 0.712 for the infarction area. However, compared to the myocardium, scar tissue segmentation still proved to be a daunting task. Methods incorporating complex pipelines or

**Table 7**
Results of the classification contest. Best results in bold.

| Challengers | Sensitivity (%) | Specificity (%) | Precision (%) | Accuracy (%) |
|---|---|---|---|---|
| Lourenço et al. | 87.88 | 70.59 | 85.29 | 82 |
| Ivantsits et al. | 72.73 | 82.35 | 88.89 | 76 |
| Sharma et al. | 72.73 | 41.18 | 70.59 | 62 |
| Girum et al.[a] | 78.79 | 88.24 | 92.86 | 82 |
| Shi et al.[a] | **90.91** | **94.12** | **96.77** | **92** |

[a] Co-authors come from the challenge organization team. Do not participate in rankings.

an important amount of parameters did not always show superiority in the results. The best segmentation approach employed two conventional U-Net variants and the configurations of nnU-Net (Isensee et al., 2021) where the first was in 2D and the second was in 3D (Zhang). All the challengers employed U-Net-based models (in 2D or 3D), and V-Net (Milletari et al., 2016) architecture could also be considered in order to exploit the 3D nature of the data. We can notice that the segmentation of the MI is a little bit different according to the considered coronary artery. Indeed, it seems that most of the proposed approaches get more difficulties to segment MI at the level of the LAD, maybe due to a more complex pathway of this artery. However, the difference between the results according to the three main coronary arteries are not significant. The best pathology classification accuracy is of 92%. This method employed an encoding CNN to predict the scar volume from the MRI, then concatenated the intermediate volume prediction to other textual features for the final classification. Considering Table 3, we could make the hypothesis that obtaining solely the volumes of the myocardium and of the cavity from a robust segmentation process is sufficient for the classification task. But there is an important overlap of the values between the two groups, rendering this approach not robust. Therefore, it could be assumed that an adaptive approach works more efficiently than attempting heavy networks. The depth of MRI and patient features' semantic information is much less than the data dedicated for human environment applications such as MS COCO and KITTI datasets (Lin et al., 2014; Geiger et al., 2013). Unless the appearance of a revolutionary new approach, a better adaptation incorporating the adequate architecture, preprocessing, training and inference etc. should be a more robust and generalized solution in the domain of medical data.

### 5.2. Inter-slice correlation

As discussed by many challengers in the segmentation contest (Feng et al., Yang and Wang, Zhang), the inter-slice information is meaningful but tricky. Anatomical facts confirm the correlation between slices, but the cardiac MRI acquisition involves anisotropic voxel size and variable numbers of slices. Such facts require challengers to weigh up the pros and cons about the use of the inter-slice correlation. The winner (Zhang) justified his trade-off such as the preliminary segmentation relied only on the intra-slice information and the final segmentation extends the receptive field to the inter-slice information. This approach reproduces the clinical practice: for most of the cases, considering a single slice is sufficient, but to distinguish the infarction and especially the PMO, the neighboring slices should be referred by physicians in case of ambiguity.

### 5.3. Gating and attention mechanism

The attention mechanism (Vaswani et al., 2017) has become a popular topic from serial data as Natural Language Processing (NLP) to computer vision tasks. The attention in neural networks mimics cognitive attention: valuable information should be enhanced and redundant information will be faded out. The attention can be applied to relatively concrete data such as the skip connections (Oktay et al., 2018), or inside a convolutional block for more abstract gating such as SE block and IRB (cf. Section 3.1.3). Unfortunately, according to the challenge results, the approaches employing the attention mechanism did not prove to outperform the vanilla U-Net or U-Net with conventional building blocks, although some challengers reported its advantage on their split validation set. An ablation study of the attention-based blocks on the state-of-the-art pipeline for the segmentation contest should be worthwhile in future work.

### 5.4. False segmentation and loss functions

Challengers, especially of the segmentation contest, have taken note on the class imbalance issue. The scar tissues represent a small number of instances in the dataset. The majority of challengers employed basically the weighted cross entropy loss, and optionally the Dice loss or generalized Dice loss (Sudre et al., 2017). The Dice loss solves the pixel-wise class imbalance problem. However, the vanilla Dice loss does not address the image-wise or the batch-wise imbalance, namely the scar tissues only exist in few images, especially the PMO. Without the weighting, the Dice loss would still suffer from the image-wise imbalance issue: the predictive model would easily assume that such targets do not exist at all, as they do in most batches. It could explain the fact that some challengers under-estimated the scar tissues if they employed the non-weighted multi-class Dice loss, in other words, the generalized Dice loss with equal class weight.

### 5.5. Data variance

Challengers investigated a variety of data augmentation methods. Such methods have been widely approved for the applications in short of training data. Nevertheless, the generated data should follow the distribution of the original data, thus completely new features should not be produced. According to this hypothesis, data augmentations such as elastic deformation and mix-up should be applied with caution. Overall, the generated features represent a fuzzy concept, only experiments can determinate if the features are bias or not. Besides the data augmentation, another approach that may increase the data variance of the training data is transfer learning. Some challengers reported the employment of transfer learning with cine MRI from the ACDC dataset (Bernard et al., 2018). The cine MRI and the DE-MRI are different acquisition techniques, but both in short axis orientation of the left ventricle. Although the challengers limited the transfer learning on only the myocardium delineation, any approach that may significantly alter the learning characteristics of the model should be undertaken with caution.

### 5.6. Clinical implications

Artificial intelligence can help make cardiovascular MRI more accessible (for both the segmentation of areas of interest and the

classification of the exams), since every clinical center has the expertise to analysis cardiovascular MRI. Specialized department relies on highly trained cardiologist to analyse and produce reports of these exams. However, these are usually limited to high volume tertiary centers. Evaluation of the presence and the extent of the myocardial infarction (with or without PMO) stays crucial in the evaluation of the myocardial viability. The visual estimation by physicians remains the routine approach, although an accurate automatic prediction of the exams as an objective evaluation of the volume and the percentage of diseased myocardium would improve the diagnosis and prognosis steps. Automatic classification allows reducing the time used to do the diagnosis and reduce the inter-expert variability. However, classification software considered as “black box” must be validated on a large and diverse dataset in order to be accepted in clinical use. Moreover, the segmentation of the different areas must be done with high accuracy and robustness. Results suggest that automatic myocardial segmentation is now a possible task, but the segmentation of diseased areas needs further development before being integrated into software solutions used in clinical practice. Moreover, in this work, only myocardial infarction is considered, and the proposed approaches must also be tested on other pathologies that involve an abnormal signal in DE-MRI, such as myocarditis or hypertrophic cardiomyopathy.

## 6. Conclusion

DE-MRI is a non-invasive technique providing the assessment of myocardial viability, but it still requires an automatic processing to get objective values of the presence and extent of the disease. In this paper, we have shown that the automatic classification of an exam between non-infarct or infarct exam is possible. Moreover, the best U-Net based methods provide an accurate delineation of the myocardium. However, the segmentation of the myocardial infarction and particularly that of the PMO area remains challenging, requiring further development to provide the extent of the infarction in a robust manner. These limitations are certainly due to the small size of the disease areas (and then due to the imbalance issue) as the lack of contrast with the surrounding structures.

## 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.

## Acknowledgment

This work was supported by the ADVANCES project founded by ISITE-BFC project (number ANR-15-IDEX-0003) and by the french ADVANCES project founded (contract ANR-17-EURE-0002).

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2022.102428.

## References

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.-M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.-M., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11), 2514–2525. doi:10.1109/TMI.2018.2837502.

Brahim, K., Qayyum, A., Lalande, A., Boucher, A., Sakly, A., Meriaudeau, F., 2021. Efficient 3D deep learning for myocardial diseases segmentation. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 359–368. doi:10.1007/2F978-3-030-68107-4_37.

Breiman, L., 1996. Bias, Variance, and Arcing Classifiers. Technical Report.

Camarasa, R., Faure, A., Crozier, T., Bos, D., de Bruijne, M., 2021. Uncertainty-based segmentation of myocardial infarction areas on cardiac MR images. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 385–391. doi:10.1007/978-3-030-68107-4_40.

Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., et al.American Heart Association Writing Group on Myocardial Segmentationfor Cardiac Imaging, R., 2002. 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 105 (4), 539–542. doi:10.1161/hc0402.102975.

Cetin, I., Sanroma, G., Petersen, S.E., Napel, S., Camara, O., Ballester, M.-A.G., Lekadir, K., 2018. A radiomics approach to computer-aided diagnosis with cardiac cine-MRI. In: Pop, M., Sermesant, M., Jodoin, P.-M., Lalande, A., Zhuang, X., Yang, G., Young, A., Bernard, O. (Eds.), Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. Springer International Publishing, Cham, pp. 82–90. doi:10.1007/978-3-319-75541-0_9.

Clevert, D., Unterthiner, T., Hochreiter, S., 2016. Fast and accurate deep network learning by exponential linear units (ELUs). In: Bengio, Y., LeCun, Y. (Eds.), 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings.

Cochet, A., Zeller, M., Cottin, Y., Robert-Valla, C., Lalande, A., L'Huilllier, I., Comte, A., Walker, P.M., Desgres, J., Wolf, J.-E., Brunotte, F., 2004. The extent of myocardial damage assessed by contrast-enhanced MRI is a major determinant of N-BNP concentration after myocardial infarction. Eur. J. Heart Fail. 6 (5), 555–560. doi:10.1016/j.ejheart.2003.11.012.

Feng, X., Kramer, C.M., Salerno, M., Meyer, C.H., 2021. Automatic scar segmentation from DE-MRI using 2Ddilated UNet with rotation-based augmentation. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 400–405. doi:10.1007/978-3-030-68107-4_42.

Fix, E., Hodges, J.L., 1989. Discriminatory analysis. Nonparametric discrimination: consistency properties. Int. Stat. Rev. 57 (3), 238–247. doi:10.2307/1403797.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: the KITTI dataset. Int. J. Robot. Res. 37, 1231–1237. doi:10.1177/0278364913491297.

Girum, K.B., Skandarani, Y., Hussain, R., Grayeli, A.B., Créhange, G., Lalande, A., 2021. Automatic myocardial infarction evaluation from delayed-enhancement cardiac MRI using deep convolutional networks. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 378–384. doi:10.1007/978-3-030-68107-4_39.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. PMLR, Fort Lauderdale, FL, USA, pp. 315–323.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 770–778. doi:10.1109/CVPR.2016.90.

Hennemuth, A., Friman, O., Huellebrand, M., Peitgen, H.-O., 2013. Mixture-model-based segmentation of myocardial delayed enhancement MRI. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 87–96. doi:10.1007/978-3-642-36961-2_11.

Hennemuth, A., Seeger, A., Friman, O., Miller, S., Oeltze, S., otto Peitgen, H., 2008. A comprehensive approach to the analysis of contrast enhanced cardiac MR images. IEEE Trans. Med. Imaging 27 (11), 1592–1610. doi:10.1109/TMI.2008.2006512.

Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18 (7), 1527–1554. doi:10.1162/neco.2006.18.7.1527.

Ho, T.K., 1995. Random decision forests. In: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1), p. 278.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. doi:10.1109/CVPR.2018.00745.

Huellebrand, M., Ivantsits, M., Zhang, H., Kohlmann, P., Kuhnigk, J.-M., Kuehne, T., Schönberg, S., Hennemuth, A., 2021. Comparison of a hybrid mixture model and a CNN for the segmentation of myocardial pathologies in delayed enhancement MRI. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 319–327. doi:10.1007/978-3-030-68107-4_32.

Isensee, F., Jaeger, P., Kohl, S., Petersen, J., Maier-Hein, K., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18 (2), 1–9. doi:10.1038/s41592-020-01008-z.

Ivantsits, M., Huellebrand, M., Kelle, S., Schönberg, S.O., Kuehne, T., Hennemuth, A., 2021. Deep-learning-based myocardial pathology detection. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 369–377. doi:10.1007/978-3-030-68107-4_38.

Karim, R., Bhagirath, P., Claus, P., James Housden, R., Chen, Z., Karimaghaloo, Z., Sohn, H.-M., Lara Rodríguez, L., Vera, S., Albà, X., Hennemuth, A., Peitgen, H.-O., Arbel, T., Gonzàlez Ballester, M.A., Frangi, A.F., Götte, M., Razavi, R., Schaeffter, T., Rhode, K., 2016. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images. Med. Image Anal. 30, 95–107. doi:10.1016/j.media.2016.01.004.

Killip, T., Kimball, J.T., 1967. Treatment of myocardial infarction in a coronary care unit: a two year experience with 250 patients. Am. J. Cardiol. 20 (4), 457–464. doi:10.1016/0002-9149(67)90023-9.

Kim, R.J., Fieno, D.S., Parrish, T.B., Harris, K., Chen, E.-L., Simonetti, O., Bundy, J., Finn, J.P., Klocke, F.J., Judd, R.M., 1999. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. Circulation 100 (19), 1992–2002. doi:10.1161/01.cir.100.19.1992.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc..

Lalande, A., Chen, Z., Decourselle, T., Qayyum, A., Pommier, T., Lorgis, L., de la Rosa, E., Cochet, A., Cottin, Y., Ginhac, D., Salomon, M., Couturier, R., Meriaudeau, F., 2020. Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. Data 5 (4). doi:10.3390/data5040089.

Li, X., Wang, W., Hu, X., Yang, J., 2019. Selective kernel networks. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019. Computer Vision Foundation / IEEE, pp. 510–519.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 740–755. doi:10.1007/978-3-319-10602-1_48.

Lourenço, A., Kerfoot, E., Grigorescu, I., Scannell, C.M., Varela, M., Correia, T.M., 2021. Automatic myocardial disease prediction from delayed-enhancement cardiac MRI and clinical information. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 334–341. doi:10.1007/978-3-030-68107-4_34.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. doi:10.1109/3DV.2016.79.

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Y Hammerla, N., Kainz, B., et al., 2018. Attention U-Net: learning where to look for the pancreas. arXiv e-prints, arXiv-1804.

Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An introduction to logistic regression analysis and reporting. J. Educ. Res. 96 (1), 3–14. doi:10.1080/00220670209598786.

Quinlan, J., 1987. Simplifying decision trees. Int. J. Man-Machine Stud. 27 (3), 221–234. doi:10.1016/S0020-7373(87)80053-6.

Ramachandran, P., Zoph, B., Le, Q. V., 2017. Searching for activation functionsarXiv:1710.05941

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Schinkel, A.F., Poldermans, D., Elhendy, A., Bax, J.J., 2007. Assessment of myocardial viability in patients with heart failure. J. Nucl. Med. 48 (7), 1135–1146. doi:10.2967/jnumed.106.038851.

Scholkopf, B., 2001. The kernel trick for distances. In: Adv. Neural Inf. Process. Syst., pp. 301–307.

Selvanayagam, J.B., Kardos, A., Francis, J.M., Wiesmann, F., Petersen, S.E., Taggart, D.P., Neubauer, S., 2004. Value of delayed-enhancement cardiovascular magnetic resonance imaging in predicting myocardial viability after surgical revascularization. Circulation 110 (12), 1535–1541. doi:10.1161/01.CIR.0000142045.22628.74.

Sharma, R., Eick, C.F., Tsekos, N.V., 2021. Sm2n2: a stacked architecture for multimodal data and its application to myocardial infarction detection. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 342–350. doi:10.1007/978-3-030-68107-4_35.

Shi, J., Chen, Z., Couturier, R., 2021. Classification of pathological cases of myocardial infarction using convolutional neural network and random forest. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 406–413. doi:10.1007/978-3-030-68107-4_43.

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support - Third International Workshop, DLMIA, pp. 240–248. doi:10.1007/978-3-319-67558-9_28.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 4278–4284.

Taghanaki, S., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G., 2019. Combo loss: handling input and output imbalance in multi-organ segmentation. Comput. Med. Imaging Graph. 75, 24–33. doi:10.1016/j.compmedimag.2019.04.005.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc..

Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. doi:10.1109/CVPR.2017.634.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.-A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C.J., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J., 2021. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Med. Image Anal. 67, 101832. doi:10.1016/j.media.2020.101832.

Yang, S., Wang, X., 2021. A hybrid network for automatic myocardial infarction segmentation in delayed enhancement-MRI. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 351–358. doi:10.1007/978-3-030-68107-4_36.

Yue, Q., Luo, X., Ye, Q., Xu, L., Zhuang, X., 2019. Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, Cham, pp. 559–567. doi:10.1007/978-3-030-32245-8_62.

Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D., 2018. mixup: Beyond empirical risk minimization. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30–May 3, 2018, Conference Track Proceedings.

Zhang, Y., 2021. Cascaded convolutional neural network for automatic myocardial infarction segmentation from delayed-enhancement cardiac MRI. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 328–333. doi:10.1007/978-3-030-68107-4_33.

Zhou, Y., Zhang, K., Luo, X., Wang, S., Zhuang, X., 2021. Anatomy prior based U-Net for pathology segmentation with attention. In: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 392–399. doi:10.1007/978-3-030-68107-4_41.