UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Deep Learning Tools for Outcome Prediction in Atrial Fibrillation from Cardiac MRI

Ana Catarina Feliciano Lourenço

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Prof. Gina Caetano
Dr. Marta Varela

2021

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Dr Teresa Matias Correia and Dr Marta Varela, who supported me every day and without whom this project would not have happened. I am especially grateful to both, for the dedication to the project and all the days and nights spent in longs meetings. It was a pleasure to grow in the personal and professional way throughout this time, and share all the victories with them. To Dr Teresa Matias Correia and Dr Marta Varela, a huge thank for their priceless guidance, help and friendship.

I would also like to express my gratitude to Professor Gina Caetano, my teacher and supervisor at Faculty of Sciences of the University of Lisbon (FCUL), for all her suggestions and advice.

I would like to acknowledge Dr Eric Keerfoot, for all the technical support, guidance and expertise, which allowed this project to be brought to success.

An acknowledgement is addressed to Erasmus for making this internship easier from a financial point of view.

I am grateful for the people I have met in London and made this internship an amazing time of my life. A huge thank to my friends that came over and brought a bit of Portugal in every visit.

Finally, a loving thank you to my family, especially my parents and my brother, for the continuous support, love and encouragement throughout my life.

RESUMO

O coração é um órgão muscular inserido na cavidade torácica e posicionado medialmente entre os pulmões. Este órgão é dividido em quatro câmaras, duas superiores (aurículas) e duas inferiores (ventrículos). O lado esquerdo (inclui a aurícula esquerdo e o ventrículo esquerdo) controla a distribuição de sangue arterial nos tecidos do corpo. Por outro lado, o lado direito (aurícula direito e ventrículo direito) é responsável pela transferência de sangue venoso do coração para os pulmões e vise-versa [1, 2]. Em particular, a **aurícula esquerda (AE)** recebe sangue arterial das veias pulmonares e bombeia-o para o ventrículo esquerdo. O AE possui ainda uma projeção em formato de dedo denominada apêndice auricular esquerdo (AAE), cuja função principal é aumentar a capacidade sanguínea do AE durante a fase de enchimento [3]. Além disso, durante o ciclo cardíaco, o AE varia o seu volume entre 35 ml e 80 ml, desempenhando diferentes funções ao longo das diferentes fases cardíacas.

Doenças cardiovasculares como enfartes do miocárdio e ritmos cardíacos anormais (arritmias), afetam o coração e/ou os vasos sanguíneos, sendo a principal causa de morte em todo o mundo [4, 5].

A fibrilhação auricular (FA) é atualmente a arritmia cardíaca mais comum, sendo caracterizada pela contração rápida e irregular das câmaras superiores do coração, as aurículas. A FA afeta mais de 33 milhões de pessoas em todo o mundo, estando associada ao aumento da mortalidade e tendo um impacto negativo na qualidade de vida dos pacientes. Embora a FA seja tratada principalmente através de medicação, a ablação por cateter é indiscutivelmente um tratamento de referência para combater diretamente esta patologia. Atualmente, as ablações são apenas eficazes no primeiro procedimento em aproximadamente 50% dos pacientes com FA [6]. A eficácia do tratamento para FA poderia ser maior se estivessem disponíveis biomarcadores capazes de caracterizar a função auricular. Em particular, as dimensões e a forma do AE têm-se demostrado preditores razoáveis do resultado das ablações [7]. Espera-se que caracterizações dinâmicas das dimensões e volumes auriculares ao longo do ciclo cardíaco sejam ainda mais informativos [6], como é o caso da fração de ejeção do AE (FE) e a fração de ejeção da contração ativa (aEF) que contêm informações clínicas importantes para controlar a FA.

CINE MRI combina boa resolução espacial e temporal com grande cobertura cardíaca e, portanto, é ideal para avaliar a função do AE. No entanto, análises temporais volumétricas detalhadas dos volumes auriculares não estão atualmente disponíveis para uso clínico por duas razões. Em primeiro lugar, as imagens CINE MRI das aurículas são tipicamente adquiridas em cortes únicos de 2 câmaras e 4 câmaras, que não mostram o volume atrial total [8]. Em segundo lugar, técnicas automáticas confiáveis para segmentar as aurículas em visualizações de eixo curto em todo o ciclo cardíaco ainda não estão disponíveis. Desta forma, na primeira fase do projecto foi proposta uma rede neural dedicada à SEGmentação totalmente automática do AE em CINE MRI de eixo curto, com base numa rede neural convolucional (SEGANet), secção 3. De forma a medir a qualidade das segmentações foram estimadas métricas como o coeficiente de Dice (CD), as distâncias de contorno de medianas (DCM) e a distância de Hausdorff (DH), entre o "ground truth" (GT ou M1) e segmentações provineantes da SEGANet Através dos resultados obtidos foi possível verificar que o método proposto produz segmentações de alta qualidade, equiparavéis ás segmentações manuais (DC: 0.93 ± 0.04, HD: 4.59 ± 2.06 mm; MCD: 0.75 ± 0.31 mm).

Em seguida, para uma visualização completa das variações de volume durante o ciclo cardíaco, foram traçadas curvas volumétricas do AE ao longo do tempo. O resultados mostraram que o volume do AE varia suavemente ao longo do ciclo cardíaco, confirmando a boa qualidade das segmentações nas diferentes fases cardíacas. A partir das curvas de tempo-volume foram automaticamente calculados o volume máximo

(V_{max}), mínimo (V_{min}) e o volume no topo da contração auricular (V_{preA}). Os volumes obtidos para o AE ao longo do ciclo cardíaco para pacientes com FA (V_{min} : 79.40 ± 25.34 mL; V_{max} : 111 ± 24.00 mL; V_{preA} : 103.40 ± 25.34 mL) e para voluntários saudáveis (V_{min} : 22.43 ± 25.34 mL; V_{max} : 44.23 ± 24.73 mL; V_{preA} : 35.47 ± 19.48 mL) demonstram grande variabilidade entre sujeitos. Além disso, a FE e a FEa foram estimados usando FE $= \frac{V_{max}-V_{min}}{V_{max}} \times 100$ e FEa $= \frac{V_{preA}-V_{min}}{V_{preA}} \times 100$. Os valores obtidos para a FE no AE (FE em Pacientes: $31.1\% \pm 9.9\%$, voluntários saudáveis: $49.8\% \pm 7.6\%$) e para a FEa (FEA em Pacientes: $24.3\% \pm 9.0\%$, voluntários saudáveis: $37.9\% \pm 10.1\%$) são, como expectável, superiores ($p < 1\text{e}^{-4}$ e $p < 1\text{e}^{-7}$, respetivamente) nos voluntários saudáveis.

A segunda parte do projeto, detalhada na secção 4, foi dedicada ao desenvolvimento de modelos de classificação baseados em imagens de ressonância magnética cardíaca. O desafio EMIDEC STACOM 2020 foi usado como um projeto inicial para criar classificadores binários, baseados em redes neurais de classificação totalmente automática. O desafio visava construir redes neurais que classificassem automaticamente a doença do miocárdio usando imagens DE-CMR e informações clínicas dos pacientes [9].

O enfarte do miocárdio (MI) ocorre quando o fluxo sanguíneo diminui devido à oclusão de uma artéria coronária, causando um enfarte (morte do tecido) no músculo cardíaco. O sintoma mais comum de MI é a dor no peito, denominada angina, que pode ou não irradiar para outras partes do corpo, como ombro ou o braço. A incidência de MI é de 32,4 milhões anualmente em todo o mundo, sendo responsável por mais de 15 % da mortalidade. O tamanho e a gravidade de um evento de MI dependem de fatos como duração e extensão do bloqueio. DE-CMR é considerada a técnica de referência não invasiva usada para avaliar o MI e a viabilidade do coração em casos de doença arterial coronariana [10, 11], podendo ainda ajudar a diferenciar doenças isquémicas do miocárdio de não isquémicas [12]. As imagens DE-CMR são tipicamente adquiridas 10-15 minutos após a injeção intravenosa de um agente de contraste à base de gadolínio. Em tecido normal o agente de contraste é expelido rapidamente, enquanto que em regiões de tecido cicatricial o processo é retardado, fazendo com que essas regiões pareçam claras em imagens T_1. O realce tardio do tecido com gadolínio no miocárdio do ventrículo esquerdo (VE) e a sua extensão, fornecem informações de diagnóstico e prognóstico importantes, como o risco de um evento cardíaco adverso e a resposta a estratégias terapêuticas, como revascularização [13]. Além disso, regiões de obstrução microvascular, apresentam uma cor escura em imagens T_1 e estão associadas a desfechos clínicos menos favoráveis .

Para esta tarefa foram criados vários modelos usando 1) apenas informação clínica dos pacientes (Clinic-NET), 2) informação clínica e imagens DE-CMR (DOC-NET). Posteriormente foi adicionada informação volumétrica para optimizar as reudes neurais: Clinic-Net+ (classificação apenas a partir de dados de texto melhorados) e DOC-NET+ (classificação a partir de dados de texto melhorados e imagens DE-CMR). Para avaliar o desempenho de cada modelo, foram calculadas métricas de qualidade num conjunto de 20 casos de teste, como a precisão, especificidade e sensibilidade. Uma vez que todos os algoritmos alcançaram bons valores de precisão (Clinic-NET +: 100 %, DOC-NET +: 100 %, DOC-NET +: 95 %, Clinic-NET: 85 %), as redes neurais foram usadas como um ponto de partida para a próxima fase do projeto.

Na última fase do projeto, os modelos de classificação anteriormente referidos foram retreinados para prever a recorrência de FA após ablação por cateter em pacientes com FA (secção 5). Usando imagens de CINE-MRI do AE na fase de pré-ablação esta tarefa obteve os melhores resultando usando a Clinic-NET + com uma precisão de teste de 88 %, seguida pela Clinic-NET (precisão: 75%) e a DOC-NET (precisão:

63%).

Em suma, o trabalho mostra o potencial das redes neurais para interpretar e extrair informações clínicas de imagens de ressonância magnética cardíaca. No futuro, se houver mais dados disponíveis, estes métodos poderão ajudar e orientar não apenas o diagnóstico, mas também o prognóstico a partir de imagens médicas.

Palavras-chave: Fibrilhação Auricular; Enfarte do miocárdio; Ressonância Magnética Cardíaca; *Deep learning*; Classificação/Segmentação.

ABSTRACT

Atrial fibrillation (AF), is the most frequent sustained cardiac arrhythmia, described by an irregular and rapid contraction of the two upper chambers of the heart (the atria). AF development is promoted and predisposed by atrial dilation, which is a consequence of atria adaptation to AF. However, it is not clear whether atrial dilation appears similarly over the cardiac cycle and how it affects ventricular volumes. Catheter ablation is arguably the AF gold standard treatment. In their current form, ablations are capable of directly terminating AF in selected patients but are only first-time effective in approximately 50% of the cases.

In the first part of this work, volumetric functional markers of the left atrium (LA) and left ventricle (LV) of AF patients were studied. More precisely, a customised convolutional neural network (CNN) was proposed to segment, across the cardiac cycle, the LA from short axis CINE MRI images acquired with full cardiac coverage in AF patients. Using the proposed automatic LA segmentation, volumetric time curves were plotted and ejection fractions (EF) were automatically calculated for both chambers.

The second part of the project was dedicated to developing classification models based on cardiac MR images. The EMIDEC STACOM 2020 challenge was used as an initial project and basis to create binary classifiers based on fully automatic classification neural networks (NNs), since it presented a relatively simple binary classification task (presence/absence of disease) and a large dataset. For the challenge, a deep learning NN was proposed to automatically classify myocardial disease from delayed enhancement cardiac MR (DE-CMR) and patient clinical information. The highest classification accuracy (100%) was achieved with Clinic-NET+, a NN that used information from images, segmentations and clinical annotations.

For the final goal of this project, the previously referred NNs were re-trained to predict AF recurrence after catheter ablation (CA) in AF patients using pre-ablation LA short axis in CINE MRI images. In this task, the best overall performance was achieved by Clinic-NET+ with a test accuracy of 88%.

This work shown the potential of NNs to interpret and extract clinical information from cardiac MRI. If more data is available, in the future, these methods can potentially be used to help and guide clinical AF prognosis and diagnosis.

Keywords: Atrial fibrillation; Myocardial disease; Deep learning; Cardiac MRI; Classification/Segmentation.

PEER-REVIEWED PUBLICATION & PRESENTATIONS AT INTERNATIONAL CONFERENCES

This thesis resulted in the following publications and presentations:

PEER-REVIEWED PUBLICATIONS

A. Lourenço, E. Kerfoot, C. Dibblin, E. Alskaf, A. Anjari, A. Bharath, H. Chubb, A. King, T. Correia and M. Varela, Automatic estimation of left atrial function from short axis CINE-MRI using machine learning, European Heart Journal, Volume 41, Issue Supplement_2, November 2020, ehaa946.0229, https://doi.org/10.1093/ehjci/ehaa946.0229.

A. Lourenço, E. Kerfoot, C. Dibblin, E. Alskaf, A. Anjari, A. Bharath, H. Chubb, A. King, T. Correia and M. Varela, Left atrial ejection fraction estimation using SEGANet for fully automated segmentation of CINE MRI, Lecture Notes in Computer Science 2020 (in press), pre-print available at https://arxiv.org/abs/2008.13718.

A. Lourenço, E. Kerfoot, I. Grigorescu, C. Scannel , M. Varela and T. Correia, Automatic Myocardial Disease Prediction From Delayed-Enhancement Cardiac MRI and Clinical Information, Lecture Notes in Computer Science 2020 (in press), pre-print available at https://arxiv.org/abs/2010.08469.

PRESENTATIONS AT INTERNATIONAL CONFERENCES

A. Lourenço, E. Kerfoot, C. Dibblin, E. Alskaf, A. Anjari, A. Bharath, H. Chubb, A. King, T. Correia and M. Varela, Automatic estimation of left atrial function from short axis CINE-MRI using machine learning, European Society of Cardiology Congress 2021, poster presentation.

A. Lourenço, E. Kerfoot, C. Dibblin, E. Alskaf, A. Anjari, A. Bharath, H. Chubb, A. King, T. Correia and M. Varela, Left atrial ejection fraction estimation using SEGANet for fully automated segmentation of CINE MRI, Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop 2020, Held in Conjunction with MICCAI 2020, 8-page paper and Oral Presentation.

A. Lourenço, E. Kerfoot, I. Grigorescu, C. Scannel , M. Varela and T. Correia, Automatic Myocardial Disease Prediction From Delayed-Enhancement Cardiac MRI and Clinical Information, Automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI (EMIDEC) 2020, Held in Conjunction with MICCAI 2020, 8-page paper and Oral Presentation.

CONTENTS

Acknowledgements i

Resumo ii

Abstract v

List of Publications and Presentations vi

List of Figures ix

List of Tables xiii

List of Abbreviations xiv

1 Introduction 1

2 Background 3

2.1 Cardiac Anatomy and Function . . . 3

2.2 Cardiovascular diseases : Atrial fibrillation (AF) and Myocardial infarction (MI) . . . 8

2.3 Cardiovascular Magnetic Resonance (CMR) . . . 10

2.3.1 Dynamic Magnetic Resonance Images (CINE-MRI) . . . 11

2.3.2 Delayed-enhancement Cardiac Magnetic Resonance (DE-CMR) . . . 12

2.4 Manual and Semi-Automatic Segmentation . . . 13

2.5 Empirical Classification . . . 14

2.6 Machine and Deep Learning . . . 15

2.6.1 Fundamentals of Deep Learning (DL) . . . 15

2.6.2 Quantitative Metrics . . . 22

2.6.3 State-of-the-art Deep Learning (DL) Methods for CMR Segmentation and Classification . . . 23

3 SEGANet 26

3.1 Methodology . . . 26

3.1.1 Data Acquisition . . . 26

3.1.2 Image Preprocessing . . . 26

3.1.3 Data Augmentation . . . 27

3.1.4 SEGANet Structure . . . 27

3.1.5 Post-processing . . . 29
3.1.6 Data Analysis . . . 29
3.2 Results . . . 29
3.3 Discussion . . . 33

4 Automatic Myocardial Disease Prediction 35
4.1 Methodology . . . 35
4.1.1 Dataset: Clinical Images and Metadata . . . 35
4.1.2 Image Preprocessing . . . 35
4.1.3 DE-CMR Automatic segmentation . . . 36
4.1.4 Data Augmentation . . . 36
4.1.5 NN for myocardial disease prediction : Clinic-NET and DOC-NET . . . 36
4.2 Results . . . 38
4.3 Discussion . . . 39

5 Classification of AF Recurrence 41
5.1 Methodology . . . 41
5.1.1 Dataset: Clinical Images and Metadata . . . 41
5.1.2 Image Preprocessing . . . 41
5.1.3 Data Augmentation . . . 42
5.1.4 NN for AF Recurrence: Clinic-NET and DOC-NET . . . 42
5.2 Results . . . 42
5.3 Discussion . . . 43

6 Conclusion 44

References 45

LIST OF FIGURES

2.1 The heart comprises two sides (left and right), separated by an interventricular septum. This structure is divided into four chambers (right atrium (RA), right ventricle (RV), left atrium (LA) and left ventricle (LV)) inside several layers (the epicardium, the thick middle myocardium, and the inner endocardium). The two atrioventricular valves (the mitral valve and the tricuspid valve) divide the atria from the ventricles. The RA includes pectinate muscles (horizontal and parallel ridges of muscle bundles) and holds structures as the superior and the inferior vena cavae. The RV is partially covered by irregular muscular components (trabeculae carneae) and is attached to the pulmonary trunk which narrows to the left pulmonary artery. The LV contains the chordae tendineae held by papillary muscles and is connected to the aorta through the aortic semilunar valve. Finally, the LA is joined to the pulmonary veins (Pvs): the left inferior pulmonary vein; the left superior pulmonary vein; the right inferior pulmonary vein; and the right superior pulmonary vein). Image available from [14]. 3

2.2 Diagram of the cardiovascular system. The blue colour represents the deoxygenated blood. In detail, the deoxygenated blood is received in the right atrium (RA) and flows into the right ventricle (RV), which pumps it through the pulmonary orifice towards the lungs (1). In the pulmonary circulation, the carbon dioxide is released, and oxygen is absorbed (2) as shown in purple. Then, oxygenated blood, presented in red, flows into the pulmonary veins (Pvs) towards the left atrium (LA). Finally, the left ventricle (LV) receives oxygenated blood from the LA and pumps it through the aortic semilunar valve into the aorta (3). In the systemic circulation (4), the blood supplies the cells with nutrients and oxygen as also described in purple. Image adapted from [15]. 4

2.3 Main phases of the cardiac cycle: **a)** Atrial and ventricular diastole. **b)** Atrial systole (AS) and ventricular diastole. **c)** Atrial diastole and ventricular systole. Image from [16]. . . . 5

2.4 Electrical Heart Pulse Propagation and the respective ECG. **a)** ECG Representation. The P wave and the QRS complex are respectively related to atrial and ventricular depolarization. The T wave describes ventricular repolarisation. **b)** Electrical Propagation of the signal through the heart. The electrical pulse is initiated by the SA node in the RA, goes through the AV node and then spreads via the His-Purkinje system reaching the ventricles. Image adapted from [17]. 6

2.5 Membrane potential in response to external electrical stimulus. Image from [18]. 6

2.6 LA and LV volume during cardiac cycle. Three phases of the cardiac cycle are indicated: end ventricular diastole (ED), end ventricular systole (ES) and atrial systole (AS). **a)** Reservoir, conduit and booster pump functions are denoted by orange, gray and blue lines respectively. **b)** The volume of blood is reduced during systole and grows during diastole. Image adapted from [19, 20]. 7

2.7 Normal heart Vs heart with AF: **a)** Normal heart in sinus rhythm and corresponding electrocardiogram (ECG). **b)** Heart during AF and corresponding ECG. Image adapted from [21]. 8

2.8 Radio frequency Catheter Ablation (RFCA) representation. The catheter deliverers the radio frequency (RF) pulses into the tissue, destroying (ablating) the tiny patches of heart cells responsible for the fibrillation. Image adapted from [22]. 9

2.9 Representation of the left ventricle (LV) wall after a myocardial infarction (MI) event. The dark red zone shows a collagen scar formed during the episode. Image from [23]. . . 10

2.10 Retrospective gating. In this approach, data are acquired throughout the whole cardiac cycle, presenting no dead spaces (when inputs are not collected). The ECG allows the estimation of the heart rate mean. Then, data points from longer and shorter RR-intervals are interpolated onto the average of the RR-interval. TR is the time interval between each acquisition. Image from [24]. 12

2.11 CINE-MRI images of a volunteer in diastole showing the **a)** 2-chamber **b)** 4-chamber **c)** short-axis (SA) view of the ventricles (where the arrow indicates the papillary muscles). Image adapted from [25]. 12

2.12 DE-CMR images acquired in all three orthogonal planes. LA = left atrium, LV = left ventricle, RV = right ventricle. **a)** 4-chamber view of an acute (48-hour) myocardial infarction (MI) with areas of microvascular obstruction (black arrowhead) surrounded by bright signals (white arrowhead) **b)** 2-chamber and **c)** short-axis (SA) views of an old MI limited by the subendocardial region of the basal inferolateral wall (arrows). Image adapted from [26]. 13

2.13 Single neuron k and a simple nonlinear model. The inputs are presented by the vector $x = (x1, x2, ..., xn)^T$; b_k denotes the bias; the synaptic weights are represented by $W_k = w_{kj}, j = 1, 2, ..., n$ where j is the index of the n inputs. $F(.)$ is the activation function applied in the group of weights inputs; y_k is the output of neuron k. Image addapted from [27]. 16

2.14 Representative structure of a CNN used in a classification task. A generic CNN learns hierarchical features by employing a stack of convolutions and pooling operations using, in this example, cardiac MR images as inputs. Then, fully connected (fc) layers reduce the spatial feature maps into a flattened vector. Image from [28]. 18

2.15 2D convolution. In this example, the upper-left element of the output tensor is formed by applying a 2 x 2 kernel to the corresponding upper-left region of the input tensor. Image from [29]. 19

2.16 U-net architecture with 32x32 pixels in the lowest resolution. While blue boxes correspond to a multi-channel feature map, the white ones represent copied feature maps. On top of the box is indicated the number of channels whereas in the lower left edge is provided the x-y size. The blue, grey, red, green and light blue arrows respectively represent 3 x 3 convolutions ReLu layers, copy and crop operations, max pooling layers, 2 x 2 upsampling convolutions and 1 x 1 convolution layers. Image from [30]. 20

2.17 Single residual block representation. x is the input vector and $F(x)$ represents the residual mapping to be learned. The operation $F + x$ is performed by a shortcut connection and element-wise addition. Image from [31]. 20

2.18 Graphic illustration of how generalization and training errors, represented by the green and dashed blue lines, respectively, depend on a network's capacity. The capacity of a model describes its complexity. The red line represents the optimal capacity, where the model best fits data. On the left end of the image, the underfitting domain is displayed where both error values are high. On the right side, the overfitting scenario shows an increase in the gap between errors. Adapted from [29]. 21

2.19 Graphic representation of underfitting, overfitting and appropriate fitting. **a)** Underfitting: the model may not capture accurately the underlying structure of the data; **b)** Optimal capacity: denotes good fitting ; **c)** Overfitting: the model function fits too closely the training data that fails on predicting accurately additional data. Image adapted from [32]. 21

2.20 **a)** Simple NN with no dropout (all neurons work and share weights among them). **b)** NN applying a regularization method, dropout. Red neurons are ignored, creating fewer connections between neurons. In subsequent training iterations, other neuronal connections will be ignored (keeping the fraction of removed connections constant) to discourage networks with a high capacity from overfitting to training data. Image adapted from [33]. 21

2.21 AtriaNet architecture. The network presented two FCN for global and local pathways. Three further convolutions were applied to merge data in the combined processing. The final output had two feature maps for every 15 x 15 patch, presenting the probability of positive/negative pixel classification. Image from [34]. 23

2.22 Overview of the method proposed by Bello et al. [35]. The autoencoder takes cardiac motion meshes as its input (RV shown in solid white, LV in red). Then, the data is flattened and dropout is applied, entering in the hidden layers. The red dot represents the central hidden layer. Finally, the data is reconstructed. 25

3.1 CMR Cine slice from the project dataset: **a)** Representative short axis slice of the ventricles with right ventricular (RV) cavity, left ventricular (LV) cavity and LV myocardium indicated by arrows; **b)** Pseudo 4-chamber comprising the atria and ventricles. The location of the short-axis (SA) slices displayed in **a)** and **c)** is also indicated ($\rightarrow$) **c)** Representative short axis slice of the atria with the left atrial (LA) body and the LA appendage (LAA) indicated by arrows. 26

3.2 Augmentation used during the training process. Representative 2D slice with: **a)** no augmentation applied; **b)** intensity scaling **c)**; additive stochastic noise; **d)** 180^{o} rotation; **e)** 2D vertical flip; **f)**Translation; **g)** K-space corruption; **h)** Smooth non-rigid deformations; **i)** Imaging shift. All the augmentation functions were implemented in slice a). 28

3.3 The proposed LA segmentation network SEGANet is based on a U-Net architecture. The segmentation network is built as a stack of layers (shown on the right), which combine the encode path of the network on the left and the decode path on the right. Subsequent layers, or the bottom connection, are denoted by "Next Layer". The bottom connection is composed of a single residual unit including two sets of convolution/normalisation/regularisation sequences. Image from [36]. 28

3.4 Training (blue curve) and validation (orange curve) loss curves for SEGANet, during 10 000 iterations. The training value was calculated at every iteration, while the validation loss was estimated at the end of every epoch. 30

3.5 LA segmentations obtained with SEGANet for one representative subject during atrial diastole, overlaid on the CINE image in: **a)** short axis, **b)** pseudo 2-chamber, and **c)** pseudo 4-chamber views. The acquired short axis CINE image stacks include both ventricles and atria. Image from [36]. 30

3.6 **a)** Short axis, **b)** pseudo 4-chamber, **c)** pseudo 2-chamber views and **d)** 3D volume rendering of the left atrium (red), right ventricle (green), left ventricle (blue) and left ventricular myocardium (yellow) automatic segmentations from one representative subject, overlaid on the CINE image. The acquired short axis CINE image stacks include both ventricles and atria. 30

3.7 **a)** Dice coefficient (DC), **b)** Hausdorff distance (HD) and **c)** mean contour distance (MCD) comparing automatic (SEGANet) and manual (M2) LA segmentations with the ground truth (GT/M1) in 13 subjects. **d)** Contours indicating the results of the manual and automatic LA segmentations in a representative slice: yellow is the SEGANet automatic segmentation, red is the manual M2 segmentation, and green is the GT/M1. Image from [36]. 31

3.8 LA volume throughout the cardiac cycle for two representative subjects: **a)** an AF patient and **c)** an healthy subject. Three LA volumes are indicated: the maximal volume (V_{max}), the minimal volume (V_{min}) and the second peak volume, corresponding to the volume at the onset of the P wave, just before active atrial contraction (V_{preA}). LA EF and aEF values are shown in each subfigure. Moreover, box plots of **b)** aEF and **d)** LA EF for AF patients and healthy subjects show that values are significantly higher in healthy subjects than in patients (LA EF p-value: $< 1e^{-4}$; aEF p-value $< 1e^{-7}$). 32

3.9 LA and LV volumes throughout the cardiac cycle for two representative subjects. Three phases of the cardiac cycle are indicated: end ventricular diastole (ED), end ventricular systole (ES) and atrial systole (AS). Image from [36]. 32

4.1 **Clinic-NET:** Classification NN based on clinical information composed with four fully connected (fc) layers. 37

4.2 **DOC-NET classification network**: a) image feature vectors, obtained from the last convolutional layer of an image feature extraction network, are concatenated with the metadata vector (M) and b) sent through four fully connected (fc) layers. Image from [9]. 38

4.3 **DE-CMR images and segmentation** of the left ventricle (LV), normal myocardium, and region of LGE uptake (if present) obtained with the proposed automatic segmentation method for two slices from two representative subjects. Image from [9]. 39

LIST OF TABLES

2.2 Confusion matrix is a table with two rows and two columns that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). . . 22

3.1 Details from the division of the 72 cases accordingly to a 0.70/0.10/0.20 ratio. 27

3.2 Hyperparameters used in SEGANet. 29

3.3 MCD obtained with networks for 3D LGE images segmentation and SEGANet used in CINE MRI image stacks. Although all the NN presented the ability to segment LA, variations in contrast and resolution comparing CINE MRI and LGE images, created different challenges for SEGANet. Consequently, direct results analysis is not ideal. . . . 33

4.1 Detailed data division of the 100 labelled cases. 36

4.2 **Confusion matrix** (actual vs predicted counts of pathological and normal cases), **accuracy, sensitivity, and specificity** obtained with all classification networks. 39

5.1 The distribution was done by randomly splitting the subjects from each class (recurrence and non-recurrence) into 6 smaller groups following the 70/10/20 ratio. Groups from distinct labels were then paired considering the same portion of cases (relative to the respective class). This division method avoids the presence of information from the same input in different sets and allows a class balance across all three groups. After division, 3 LA slices from each case were processed and repeatedly concatenated to clinical information regarding the sample. 41

5.2 **Confusion matrix** (actual vs predicted counts of AF recurrence and non-recurrence cases), **accuracy, sensitivity, and specificity** obtained with all re-trained classification networks. DOC-NET accepted 3 LA slices per patient, using 12 inputs in the test set. Since CLINIC-NET and CLINIC-NET+ used one input tensor per subject, both test sets only comprised 8 cases. 43

LIST OF ABBREVIATIONS

aEF Atrial Contraction Ejection Fraction

AF Atrial Fibrillation

AI Artificial Intelligence

AP Action Potential

AS Atrial Systole

AV node Atrioventricular node

BMI Body Mass Index

bmp Beats Per Minute

bSSFP Balanced Steady-state Free Precession

CA Catheter Ablation

CE Cross-Entropy

CINE-MRI Dynamic Magnetic Resonance Image

CMR Cardiovascular Magnetic Resonance

CNN Convolutional Neural Network

CVD Cardiovascular Disease

DC Dice Coefficient

DE-CMR Delayed-enhancement Cardiac Magnetic Resonance

DL Deep Learning

ECG Electrocardiogram

ED End Ventricular Diastole

EF Ejection Fraction

ES End Ventricular Systole

FC Fully Connected

FCN Fully Connected Network

FN False Negatives

FP False Positives

FT Fourier Transform

GT Ground Truth

HAAM Hierarchical Aggregation Module

HAANet Attention Based Hierarchical Aggregation Network

HD Hausdorff Distance

LA Left Atrium

LAA Left Atrial Appendage

LAGO Left Atrial Geometry and Outcome

LGE Late Gadolinium Enhancement

LIPV Left Inferior Pulmonary Vein

LSPV Left Superior Pulmonary Vein

LV Left Ventricle

MCD Median Contour Distance

MI Myocardial Infarction

MKL Multiple Kernel Learning

ML Machine Learning

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

NN Neural Network

NSTEMI Non-ST Elevation Myocardial Infarction

preA Onset of Active Atrial Contraction

PReLU Parametric Rectified Linear Unit

PSIR Phase Sensitive Inversion Recovery

Pv Pulmonary Vein

PVI Pulmonary Vein Isolation

RA Right Atrium

ReLU Rectified Linear Unit Activation Function

Res-Net Residual Network

RF Radio Frequency

RFCA Radio Frequency Catheter Ablation

RIPV Right Inferior Pulmonary Vein

ROI Region of Interest

RSPV Right Superior Pulmonary Vein

RV Right Ventricle

SA Short-Axis

SA node Sinoatrial node

SEGANet Neural network for fully automatic SEGmentation of the left Atrium

STEMI ST-elevation Myocardial Infarction

TN True Negatives

TP True Positives

V-FCNN Volumetric Fully Convolution Neural Network

1 INTRODUCTION

Atrial fibrillation (AF) is the most frequent sustained cardiac arrhythmia, affecting more than 33 million people all over the world. It is described by an irregular and rapid contraction of the two upper chambers of the heart (the atria), when the electrical activation is not under the control of the sinoatrial (SA) node. AF impacts the patients' quality of life and is associated with increased morbidity and mortality. Despite essentially controlled with medical therapy, catheter ablation (CA) is arguably the AF gold standard treatment. In their current form, ablations are capable of directly terminating AF in selected patients but are only first-time effective in approximately 50% of the cases [6].

Biomarkers able of characterising atrial function and predict treatment outcome could highly improve AF treatment efficacy, if they were available. In particular, left Atrium (LA) dimensions and shape have revealed to be reasonable predictors of post-ablation patient outcome [7]. Thus, dynamic characterisations of atrial dimensions and volumes across the whole cardiac cycle are expected to be highly informative [37] and could be obtained from dynamic (CINE) Magnetic Resonance Images (MRI). CINE MRI is a non-invasive imaging technology able to visualise the cardiovascular anatomy and function. This technique can be used to identify wall motion abnormalities throughout the cardiac cycle and calculate functional parameters, such as heart chambers' volumes [38].

In particular, LA ejection fraction (EF) and active atrial contraction ejection fraction (aEF) comprise critical clinical information which can contribute to manage AF. The LA EF describes the LA global function across the cardiac cycle, while the aEF evaluates the LA pump function, which is expected to give additional clues about atrial tissue health. Both biomarkers are estimated employing ratios of LA volumes across the cardiac cycle. LA volumes are currently calculated with standard formulae, using assumptions on the LA shape and length information from single-slice CINE MRI or 2D atrial echocardiography [39, 40]. Moreover, aEF relies on the identification of the onset of active atrial contraction (preA), which is hard to detect visually, even for a specialist, in a single 2D view [39–41]. Hence, standard techniques allow automatic EF estimation but not aEF. Therefore, more accurate and precise results should be obtained by estimating these biomarkers from 3D CINE MRI images.

Detailed analyses of atrial volumetric temporal function and post-ablation outcome predictions are not currently available for clinical use, for two key reasons. First, atrial CINE MRI images typically only sample the ventricles, not covering the full atrial volume or are usually obtained on single-slice 2-chamber and 4-chamber views [8]. Thus, accurate automatic techniques for the segmentation of ventricular structures from short axis CINE MRI and the LA in 2- and 4-chamber views [28, 42] are available, but not for atrial segmentation in short axis CINE MRI views. Multi-slice CINE MRI is ideally suited for evaluating LA function since it combines high spatial and temporal resolution with large cardiac coverage.

In this work, it is first proposed, and described in section 3, a dedicated neural network (NN) for fully automatic SEGmentation of the LA in short axis CINE MRI based on a convolutional neural Network (SEGANet). To evaluate the performance of SEGANet, metrics such as the Dice coefficient (DC), median contour distances (MCD) and Hausdorff distances (HD) were used. Additional comparisons between SEGANet results and inter-observer variability of manual segmentations are also presented. Then, using the SEGANet LA segmentations, LA volumes across all cardiac phases and LA EFs and aEFs (in both healthy subjects and AF patients) were automatically assessed [43].

The second part of the project, detailed in section 4, was dedicated to developing classification models based on cardiac MRI images. The EMIDEC STACOM 2020 challenge was used as an initial project and basis to create binary classifiers based on fully automatic classification NNs. For this challenge, a deep

learning (DL)-NN was proposed to automatically classify myocardial disease from delayed enhancement cardiac MR (DE-CMR) and patient clinical information [9].

DE-CMR is known as the non-invasive gold standard for evaluating viability in coronary artery disease and myocardial infarction (MI) [11]. This technique can also help distinguish ischemic from non-ischemic myocardial disorders [12]. The underlying concept of this approach is based on the delayed washout of a gadolinium-based contrast agent from injured myocardium areas, making them appear bright in T_1-weighted images. However, patients with certain cardiomyopathies (such as extensive hibernating myocardium, hypertrophic cardiomyopathy, sarcoidosis or myocarditis) and no-reflow regions, may not show contrast uptake and hence, LGE absence in T_1-weighted images does not exclude the probability of myocardial disease presence [12,44,45], making disease classification from DE-CMR a hard task. Several machine learning (ML) classification models, like support vector machines [46], random forests [47], and K-nearest neighbour [48], have been used to address this important clinical gap. However, these methods demand high domain knowledge to create good inputs for the classifier. Also, complicated feature extraction procedures are required to obtain reliable results leading to a poor generalization capability.

DL architectures are able to learn features directly from the data, reducing the hard feature extraction process and expertise needed [49]. Thus, for the EMIDEC STACOM 2020 challenge, deep NNs were proposed to predict myocardial disease vs normal cases considering: 1) patient clinical information only (Clinic-NET), 2) clinical information and DE-CMR images (DOC-NET). Since cardiomyopathies can cause disturbances in LV volume and wall thickness, volumes across the DE-CMR images of relevant cardiac regions (LV myocardium, LV blood pool, LGE uptake region and no-reflow area) were estimated. This volumetric information was assessed from ventricular segmentations and then used as extra metadata in both networks: Clinic-Net+ (classification from enhanced metadata only) and DOC-NET+ (prediction from enhanced metadata and DE-CMR images).

To evaluate the performance of each network, quality metrics, such as accuracy, specificity and sensitivity, were calculated on a 20-case test dataset. Since all approaches achieved high test accuracies in the challenge (Clinic-NET+: 100%, DOC-NET+: 100%, DOC-NET+: 95%, Clinic-NET: 85%), these networks were used as a starting point for the next task of this research project.

For the final aim of the project, the previously referred classification NNs were re-trained to predict AF recurrence after CA in AF patients using pre-ablation LA short axis in CINE MRI images (section 5). In this task, the best overall performance was achieved by Clinic-NET+ with a test accuracy of 88% followed by Clinic-NET (accuracy: 75%) and DOC-NET (accuracy: 63%).

All this work showed the potential of NNs to interpret and extract clinical information from cardiac MRI. In the future, as more data becomes available, these methods will be able to help and guide clinical AF prognosis and diagnosis.

2 BACKGROUND

2.1 CARDIAC ANATOMY AND FUNCTION

The heart is a muscular organ inside the thoracic cavity, medially positioned between the lungs and surrounded by a double layer membrane, called the pericardium. This vital organ sits in its own space, the pericardial cavity, and is divided into four chambers as shown in figure 2.1: two upper chambers (atria) and two lower chambers (ventricles). The left side (comprising the left atrium (LA) and the left ventricle (LV)) controls the distribution of oxygen-rich blood into the body tissues, as described in figure 2.2. The right side (right atrium (RA) and right ventricle (RV)) is responsible for the transfer of oxygen-poor blood from the heart to the lungs for blood oxygenation [1, 2].

The average dimension of the human heart resembles a man's closed fist, reaching 15.5cm of transverse diameter in males and 14.5cm in females. The LA particularly has approximately 2mm of thickness, a normal diameter lower than 4.1cm in men and smaller than 3.9cm in women [50–52]. Compared with LA dimensions, the LV presents average values considerably higher for both men (normal size: 50.5 mm, thickness: 7 mm) and women (normal size: 46 mm, thickness: 6 mm) [53].

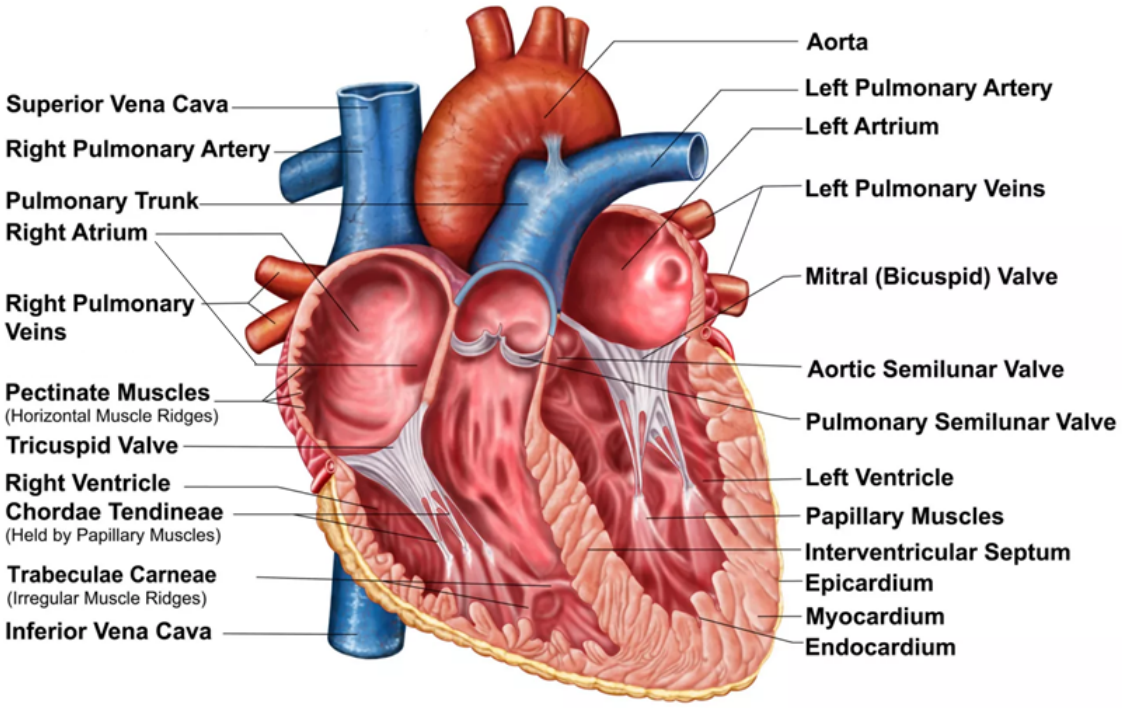

Figure 2.1: The heart comprises two sides (left and right), separated by an interventricular septum. This structure is divided into four chambers (right atrium (RA), right ventricle (RV), left atrium (LA) and left ventricle (LV)) inside several layers (the epicardium, the thick middle myocardium, and the inner endocardium). The two atrioventricular valves (the mitral valve and the tricuspid valve) divide the atria from the ventricles. The RA includes pectinate muscles (horizontal and parallel ridges of muscle bundles) and holds structures as the superior and the inferior vena cavae. The RV is partially covered by irregular muscular components (trabeculae carneae) and is attached to the pulmonary trunk which narrows to the left pulmonary artery. The LV contains the chordae tendineae held by papillary muscles and is connected to the aorta through the aortic semilunar valve. Finally, the LA is joined to the pulmonary veins (Pvs): the left inferior pulmonary vein; the left superior pulmonary vein; the right inferior pulmonary vein; and the right superior pulmonary vein). Image available from [14].

The **right atrium (RA)** forms the right border of the heart, and it is internally divided into two parts. The smooth posterior portion holds most of the connected structures, such as superior and inferior vena cavae (figure 2.1) and the natural pacemaker of the heart, named sinoatrial (SA) node. The anterior region

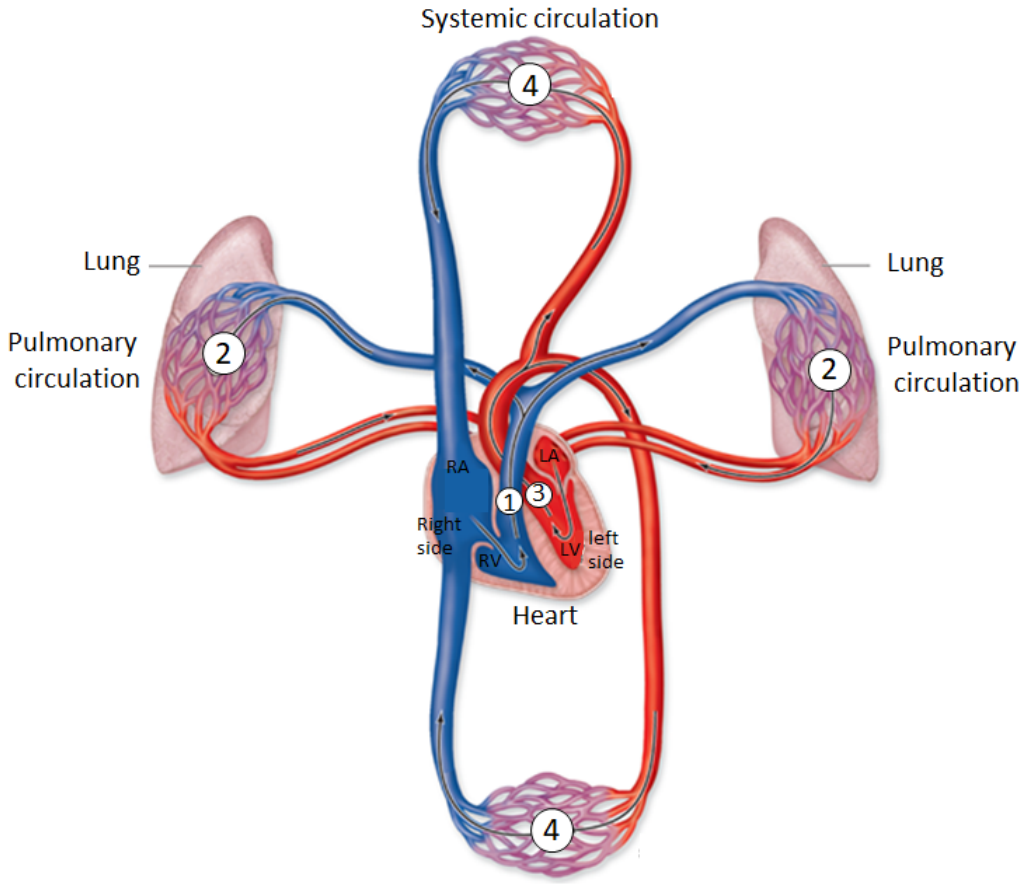

Figure 2.2: Diagram of the cardiovascular system. The blue colour represents the deoxygenated blood. In detail, the deoxygenated blood is received in the right atrium (RA) and flows into the right ventricle (RV), which pumps it through the pulmonary orifice towards the lungs (1). In the pulmonary circulation, the carbon dioxide is released, and oxygen is absorbed (2) as shown in purple. Then, oxygenated blood, presented in red, flows into the pulmonary veins (Pvs) towards the left atrium (LA). Finally, the left ventricle (LV) receives oxygenated blood from the LA and pumps it through the aortic semilunar valve into the aorta (3). In the systemic circulation (4), the blood supplies the cells with nutrients and oxygen as also described in purple. Image adapted from [15].

.

includes pectinate muscles, horizontal and parallel ridges of muscle bundles [54] also shown in figure 2.1.

The **left atrium (LA)** receives oxygenated blood from the Pvs and pumps it through the left atrioventricular orifice (which is controlled by the mitral valve) into the LV (figure 2.2). The oxygenated blood comes from the lungs, flows firstly into venules and later in the four main Pvs (described in figure 2.1): the left inferior pulmonary vein; the left superior pulmonary vein; the right inferior pulmonary vein; the right superior pulmonary vein. The Pvs are inserted in LA through four ostia arranged into two pairs. Since orifices from the same side can merge, the number of insertions is highly variable: 4 is the usual amount but some people also have 3 or 5. Regarding the interior wall, this can be divided into two parts. The smooth inflow portion receives blood from the Pvs, while the outflow portion is lined by pectinate muscles to increase the force of contraction. The LA has a finger-like projection called left atrial appendage (LAA), whose the main function is to increase the LA blood capacity during the filling phase. Usually, the LAA extends from the main body of the LA, between the anterior and the lateral walls, and has an oval orifice. Its tip, directed anterosuperiorly, overlaps some of the structures such as, the left border of the RV outflow tract and the pulmonary trunk. The LAA has considerable variations between hearts, namely in its flattened shape, size and contacts with structures, which can be extremely relevant in

surgeries. Thus, it is common to find a triangular or water-drop tips directed laterally and backwards [3].

The **right ventricle (RV)** forms the largest portion of the anterior border of the heart, receiving deoxygenated blood from the RA and pumping it through the pulmonary orifice towards to the pulmonary trunk (figure 2.2). This muscle can be divided into two regions: the outflow and the inflow tract. The outflow region, also named conus arteriosus (infundibulum), leads to the pulmonary artery and has smooth walls. The inflow tract is covered by irregular muscular structures called trabeculae carneae (figure 2.1), crucial for contraction.

The **left ventricle (LV)** receives oxygenated blood (figure 2.2) from the LA and pumps it through the aortic semilunar valve into the aorta. The LV muscular layer is also divided into an inflow and an outflow portion. The inflow region contains papillary muscles (figure 2.1), whose function is to shortly contract before the ventricular systole and then maintain the tension, preventing regurgitation. The chordae tendineae, also shown in figure 2.1, connect the mitral valve to the papillary muscles. The outflow tract is a smooth-walled also known as aortic vestibule [1, 55].

The heart is composed with several layers, the epicardium, the thick middle myocardium and the inner endocardium, shown at the right button of figure 2.1. In particular, myocardium is the main contracting layer of the heart, containing circular and spiral arrangements of muscular fibres. Cellular components include cardiac muscle cells (cardiomyocytes), contractile units of the layer crucial to produce mechanical tension. The cardiomyocytes communicate via specialised cell-to-cell junctions (intercalated disks), which allow ions to flow across adjacent myocytes, for rapid conduction of action potentials (APs) and electrical propagation of signals across the heart [2].

CARDIAC CYCLE

At the very beginning of the cardiac cycle, in the early phase of ventricular diastole, all the chambers are relaxed (atrial and ventricular diastole), allowing the veins to empty blood into the atria as shown in figure 2.3a). Since the mitral and tricuspid valves are open, the ventricles are filled up with blood from the atria. Then, during late ventricular diastole, both atria start to contract (atrial systole, AS) augmenting the ventricular filling (figure 2.3b)). Afterwards, the atrio-ventricular valves close and the pulmonary semilunar valves open, as the ventricles contract (ventricular systole), pumping the blood into circulation. At the same time, the atria are relaxed (atrial diastole) and act as reservoirs for pulmonary venous return (figure 2.3c)).

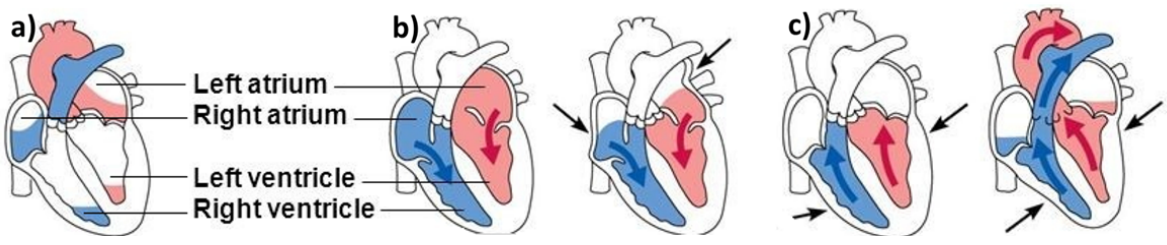

Figure 2.3: Main phases of the cardiac cycle: **a)** Atrial and ventricular diastole. **b)** Atrial systole (AS) and ventricular diastole. **c)** Atrial diastole and ventricular systole. Image from [16].

During sinus rhythm, electrical pulses generated in the SA node are propagated across the heart, triggering the coordinated rhythmic contraction of the four chambers. The electrical heart pulses can be detected by placing electrodes on the surface of the torso, in an electrocardiogram (ECG), as presented in figure 2.4a). The primary electrical pulse is initiated by the SA node as shown in figure 2.4b), within a

rate between 60 and 100 beats per minute (bmp). The SA node, acts as a natural pacemaker, comprising a collection of specialised cells, autonomous electrical oscillators whose membrane electrical potential spontaneously increases (depolarises) and decreases (repolarises) periodically. These electrical impulses are known as APs (action potentials) and described in the next paragraph. The AP first depolarise the atria, forming the P wave. Then, the AP reaches the ventricles by the atrioventricular node (AV node) and spreads through them via the His-Purkinje system, a specialised conducting tissue also presented in figure 2.4b). The rapid and coordinated depolarisation (followed by a contraction) of the ventricles produces the QRS complex. Finally, from the ventricular repolarisation arises a T wave [56].

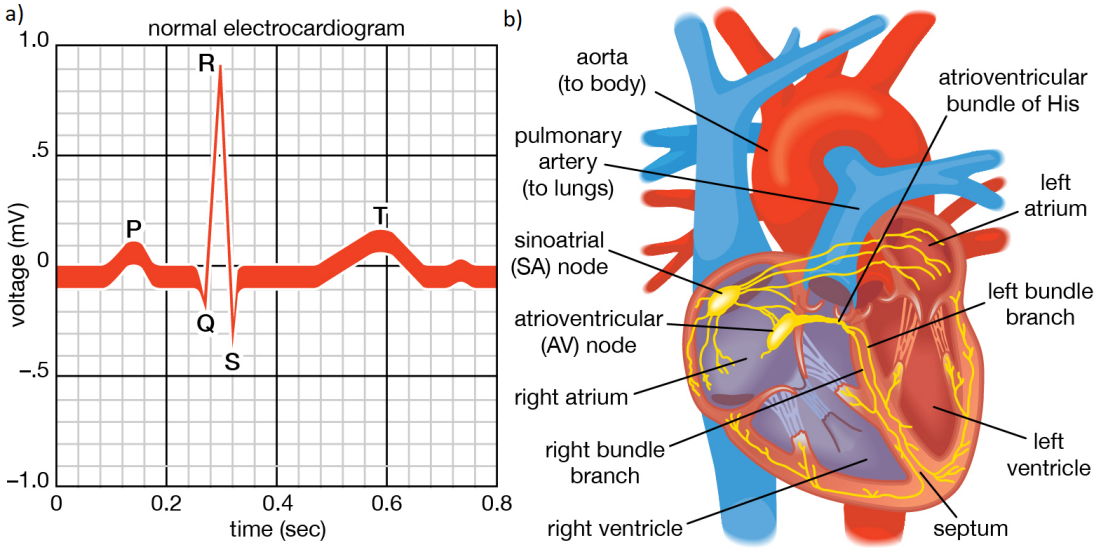

Figure 2.4: Electrical Heart Pulse Propagation and the respective ECG. **a)** ECG Representation. The P wave and the QRS complex are respectively related to atrial and ventricular depolarization. The T wave describes ventricular repolarisation. **b)** Electrical Propagation of the signal through the heart. The electrical pulse is initiated by the SA node in the RA, goes through the AV node and then spreads via the His-Purkinje system reaching the ventricles. Image adapted from [17].

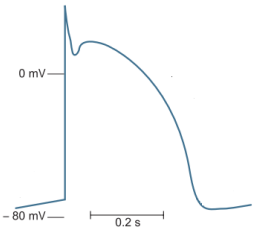

Figure 2.5: Membrane potential in response to external electrical stimulus. Image from [18].

The atrial AP, figure 2.5, is a non-linear variation in the membrane potential of atrium cells in response to external stimulus. The AP starts with a large inward (depolarizing) current of Na^+ (I_{Na}) that takes the cell from the resting potential (-85 to -65 mV) to approximately +40 mV. Then, the I_{Na} is inactivated and outward K^+ currents partially repolarise the cell causing a dent in the AP. The next phase is characterized by a plateau in the AP maintained by the balance of the K^+ currents and an inward Ca^{2+} current (I_{Ca}). Moreover, the Ca^{2+} inflow increases its concentration in the cell and produces the myocyte contraction when reaching high values. Finally, additional K^+ channels are progressively activated, creating outward currents and fully repolarizing the cell [18].

CARDIAC FUNCTIONAL PARAMETERS

The volume of the heart chambers varies through the cardiac cycle. In particular, the LA performs different roles and its volume changes between 35 mL and 80 mL, as described in figure 2.6a). During the conduit function (gray line in figure 2.6a)), the LA transfers blood passively into the LV being influenced by atrial compliance. The booster pump function (blue curve in figure 2.6a)), augments the LV filling at the AS reflecting the magnitude of atrial contractility. When the mitral valve closes the LA obtains the minimum volume. The reservoir function represented in orange at figure 2.6a), is played in the ventricular systole where LA is filled up. Just before mitral valve opening the LA reaches the maximum volume. The LV volume also varies throughout the cardiac phases (figure 2.6b)). In particular, it decreases between the end ventricular diastole (ED) and the end ventricular systole (ES), describing the blood transfer into circulation (systole). Then, during diastole the LV is filled up with blood again increasing its volume.

The global function or EF of both chambers can be estimated as :

$$\text{EF} = \frac{V_{max} - V_{min}}{V_{max}} \times 100 \tag{2.1}$$

In detail, the LA aEF evaluates the LA pump function giving additional clues about atrial tissue health, and be calculated as:

$$\text{aEF} = \frac{V_{preA} - V_{min}}{V_{preA}} \times 100 \tag{2.2}$$

where the maximal (V_{max}) and minimal (V_{min}) LA/LV volumes, as well as the LA volume at the onset of atrial contraction (V_{preA}) can be automatically calculated from the volumetric time graphs [19].

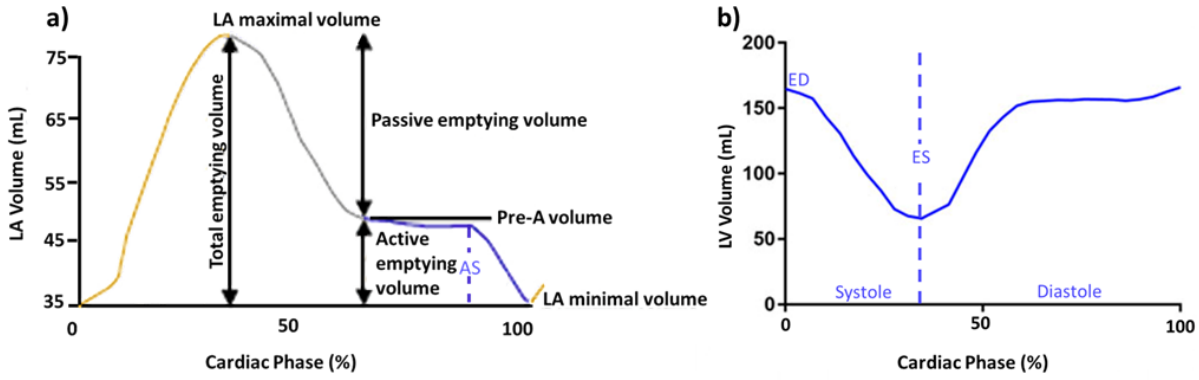

Figure 2.6: LA and LV volume during cardiac cycle. Three phases of the cardiac cycle are indicated: end ventricular diastole (ED), end ventricular systole (ES) and atrial systole (AS). **a)** Reservoir, conduit and booster pump functions are denoted by orange, gray and blue lines respectively. **b)** The volume of blood is reduced during systole and grows during diastole. Image adapted from [19, 20].

2.2 CARDIOVASCULAR DISEASES : ATRIAL FIBRILLATION (AF) AND MYOCARDIAL INFARCTION (MI)

Cardiovascular diseases (CVDs), such as MIs and abnormal heart rhythms (arrhythmias), affect the heart or blood vessels and are the number 1 cause of death worldwide, being responsible for an estimated 31% of all deaths [4, 5].

Atrial fibrillation (AF) is a type of arrhythmia, characterised by rapid and irregular activations of the atria, achieving values as high as 400–600 bmp. AF occurs when the electrical activation of the atria is not under the control of the SA node, creating rapid and uncoordinated activations of different areas in the upper chambers of the heart. Thus, this irregular activation makes the atrium quiver instead of contracting coordinately and generates an abnormal ECG as the one represented in figure 2.7b) [18].

AF affected 32.5 million people worldwide in 2010 and by 2030 its incidence in Western Europe is expected to rise to 3% of all adults aged over 20 [57]. Regarding the hazards, AF is known as the most important cause of ischaemic stroke in people older than 75; it is also related to a higher probability of heart failure and can reduce the patients' quality of life, often leading to severe disabling symptoms [7, 18].

The atria can adapt to being in AF (remodelling), altering its electrical and structural properties. The remodelling process induces the atrial dilatation predisposing and promoting the development of AF ("AF begets AF"). Through dilatation, the atrium becomes larger. However, it is not clear whether these size and shape changes are similar across the cardiac cycle. Apart from dilatation, structural remodelling includes an increase in collagen deposition around cardiomyocytes (fibrosis), which can enlarge myocyte-myocyte distances and lead to failure in the conduction of AP. Furthermore, during the electrical remodelling process, atrial electrophysiological properties are changed, leading to a shortening of the AP duration. Thus, the refractory period (time following an AP when it is more difficult to depolarise the refereed cell) decreases and the probability of AF occurrence grows [18, 58].

The disease starts as paroxysmal or intermittent AF, with less than seven days episodes that can terminate spontaneously. Due to the above remodelling processes, AF tends to get worse across time, reaching episodes with more than a week (persistent AF) and can become permanent when medical interventions can no longer restore sinus rhythm.

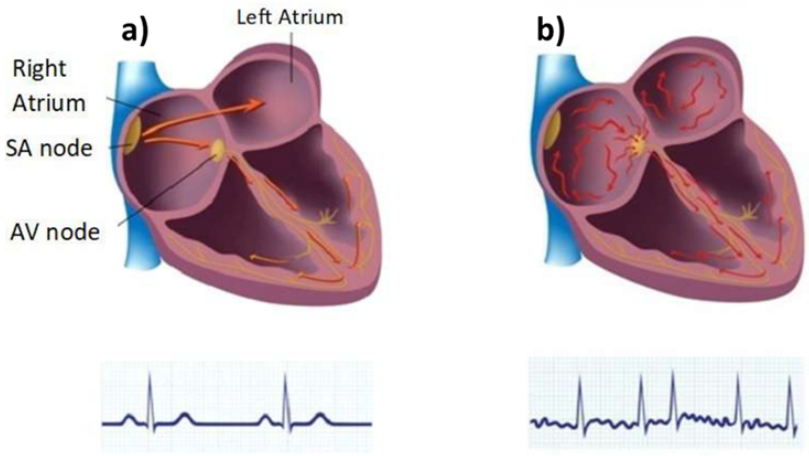

Figure 2.7: Normal heart Vs heart with AF: **a)** Normal heart in sinus rhythm and corresponding electrocardiogram (ECG). **b)** Heart during AF and corresponding ECG. Image adapted from [21].

There are three main treatments for AF: 1) anti-arrhythmic drugs, that control the cardiac rhythm by changing the electrical properties of the heart; 2) cardioversion devices which defibrillate the atria to

stop the arrhythmias with controlled electrical discharges; and 3) CA therapy. During a radio frequency catheter ablation (RFCA) the abnormal electrical activity areas are destroyed by localised high-energy pulses delivered by catheters. There is no well-defined guideline for this procedure, but the creation of lesions that completely surround the Pvs (pulmonary vein isolation, PVI), is known as the best strategy for catheter ablations (CAs). Despite data indicating better sinus rhythm maintenance after CAs than other therapies, 47% of all patients experience AF recurrence within 3 years of the first procedure and require additional ablations [6]. Furthermore, CAs are an invasive and costly procedure, which can lead to rare but life-threatening complications such as tamponade or oesophageal fistulae [59]. It is therefore paramount to carefully only select patients who are likely to benefit from ablations to undergo such procedure. Several demographic, clinical and electrophysiological biomarkers have been considered for patient stratification, ranging from patient age, duration of AF, presence of comorbidities or properties of the electrical signal detected in the atria, as well as linear combinations thereof [57, 59]. Disappointingly, all current biomarkers for post-ablation AF recurrence have shown a poor predictive value [57].

Several imaging features have been proposed to capture the features of AF-induced structural remodelling that hamper the effectiveness of CAs. These proposed biomarkers range from increases in the left atrial volume [60] or its surrogates to alterations in the spherical shape of the left atrium (LA) [61] or, in the symmetry of the LA shape [7]. Recent studies have additionally measured atrial deformation parameters to predict AF recurrence post-ablation. Using echocardiographic techniques, metrics of strain and strain rate in the LA have been found to have some predictive power for freedom from AF following CAs [62]. However, these techniques do not allow a full characterisation of atrial biomechanics, as they traditionally cover a limited portion of the LA. These only allow in-plane deformations to be estimated and suffer from high noise levels, particularly in the distal portions of the atria [63].

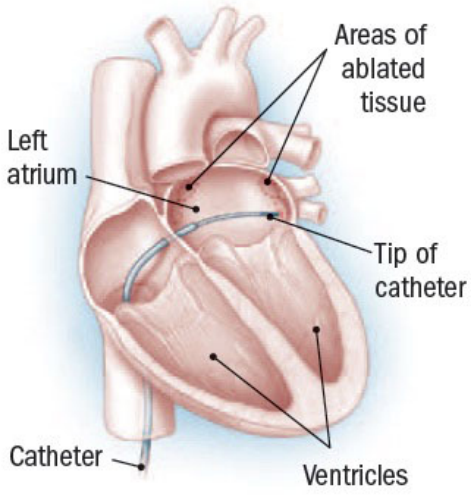

Figure 2.8: Radio frequency Catheter Ablation (RFCA) representation. The catheter deliverers the radio frequency (RF) pulses into the tissue, destroying (ablating) the tiny patches of heart cells responsible for the fibrillation. Image adapted from [22].

Myocardial infarction (MI) occurs when the blood flow decreases due to occlusion of a coronary artery, causing an infarction (tissue death) in the heart muscle. The most common symptom of MI is chest pain, called angina, that may or may not radiate to other parts of the body, such as the shoulder and the arm. The incidence of MI is 32.4 million every year in all over the world, leading to over 15% of mortality.

The main cause of MI is the rupture of atherosclerotic plaques on arteries that supply the heart.

Atherosclerosis is a progressive inflammatory disorder of the arterial wall characterised by continuous deposits of atheroma. These plaques remain clinically silent until they reach sizes able to cause tissue hypoperfusion and myocardial ischemia, or disrupt the lesion resulting in thrombotic occlusion of the vessel.

MI is commonly clinically classified into ST-elevation MI (STEMI) and non-ST elevation MI (NSTEMI). While STEMI is recognized by a typical elevation in the "ST segment" of ECG (Figure 2.4a)), the NSTEMI does not show modifications in ECG but presents positive cardiac biomarkers as troponin [64]. The treatment depends on the MI type and aims to preserve not only the heart muscle but also prevent further complications. The immediate aid uses aspirin medication, and in STEMI cases, reperfusion therapy is then adopted to restore blood flow. Reperfusion therapy includes drugs as thrombolytics, to deal with blocked arteries, and fibrinolytics to prevent blood clots from growing. Minimally-invasive endovascular procedures such as percutaneous coronary intervention are also approached [65].

The size and severity of a MI event depends on facts such as blockage duration and extension. All cells can recover if reperfusion happens after a short period of ischemia. However, as the duration of ischemia increases, cells stay longer in hypoxia, triggering necrosis, apoptosis and cardiac muscle becomes irreversibly injured. The number of dead cells grows over time, leading to form a collagen scar (figure 2.9). Tissue death and scar formation change the natural conduction on heart. Moreover, depending on the event size and location, there is a risk of arrhythmias, aneurysm of the heart ventricles and inflammation of the heart wall.

Myocardium injuries also occur during reperfusion and include tunned myocardium, reperfusion-induced arrhythmias, no-reflow phenomenon (microvascular obstruction) and lethal myocardial cell injury. In particular, the no-reflow phenomenon describes the incapacity to perfuse myocardium portions due to microvascular obstruction, despite successful artery reopening [66]. Microvascular dysfunction is a consequence of Ca^{2+} accumulation and oxygen radicals discharge. During ischemia, the non-function of ion pumps raises the concentration of Ca^{2+} into the cells. Then, in reperfusion, when oxygen is reintroduced, the quantity of Ca^{2+} is further increased due to I_{Ca} in the AP, and damages into the electron transport chain leads to oxygen radicals accumulation. The modification in ion homeostasis results in cell death and microvascular obstruction. The no-reflow areas typically observed into MI zones, myocyte cell death, and the size is correlated with the duration of coronary occlusion [67].

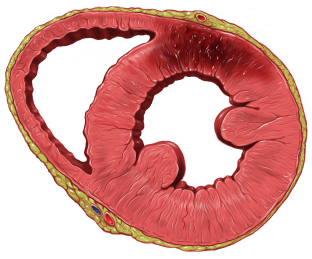

Figure 2.9: Representation of the left ventricle (LV) wall after a myocardial infarction (MI) event. The dark red zone shows a collagen scar formed during the episode. Image from [23].

2.3 CARDIOVASCULAR MAGNETIC RESONANCE (CMR)

Cardiovascular magnetic resonance imaging (CMR) is a non-invasive imaging technology, with the ability to visualise the cardiovascular anatomy and function in a single examination, being perfect to characterise the heart contractile function. Thus, this approach can be used to identify wall motion

abnormalities throughout the cardiac cycle and calculate functional parameters, such as ventricular volumes.

CMR adopts the same basis as other magnetic resonance imaging (MRI) techniques. MRI uses magnetic properties of hydrogen nuclei in water molecules of the human body to generate an image. Very briefly, in the presence of an external magnetic field, the magnetization of hydrogen nuclei (protons) tends to rotate (precess) around the external field. When a resonant radio frequency (RF) pulse is applied, the orientation of the magnetization of these protons changes in a coherent manner. When the RF pulse is then removed, the protons tend to both return to their resting state (longitudinal relaxation) and lose coherence (transverse relaxation), emitting a signal used to produce the image. The time scales of the longitudinal and transverse relaxation processes (typically measured using T_1 and T_2 values, respectively) vary across different body tissues and are used as the main source of image contrast in MRI [68].

The data are saved in k-space, where each point represents the contribution of a single spatial frequency to the whole image. Low spatial frequencies are located in the centre of k-space l, whereas the edges contain the contribution of high spatial frequencies. The 2D or 3D Fourier transform (FT) can decompose the k-space data from the frequency domain into the image space, where the spatial location are correctly represented [24].

MRI is a technically challenging imaging approach. One of the current main goals is to speed up the inherently slow data acquisition without compromising the required high resolution and image quality. Limitations in imaging moving organs occur due to the slow process. Motion is the principal source of image degradation, causing characteristic artefacts, blurring and misregistration. In particular, in Dynamic Magnetic Resonance Images (CINE-MRI), dealing with movement caused by heart beating and patient breathing persists one of the main challenges.

The motion induced by the heart pumping chambers is complex, and in current clinical protocols, data is usually obtained along with several cardiac cycles. For this reason, CMR techniques are synchronized with ECGs. The R-wave (figure 2.5a)) can be detected by the scanner and used to trigger the image acquisition, thereby reducing the cardiac motion artefacts [69, 70].

CMR images are typically acquired in one of three orthogonal planes. The long axis is defined as the line between the centre of the mitral valve and heart apex. The short-axis (SA) views are images slices perpendicular to the long axis. The 4-chamber and 2-chamber planes are orthogonal to one another and contain the long axis [71]. Moreover, CMR presents several imaging types depending on applications. In particular, CINE-MRI, shown in figure 2.11, provide complete information regarding heart function during the cardiac cycle. Also, delayed-enhancement cardiac magnetic resonance (DE-CMR) uses a contrast agent to characterize and identify potential lesions in cardiac tissue.

2.3.1 DYNAMIC MAGNETIC RESONANCE IMAGES (CINE-MRI)

CINE-MRI (figure 2.11) is is the gold-standard for the assessment of cardiac function. CINE can be especially useful to estimate metrics such as stroke volume, EF, end-diastolic and end-systolic volumes and masses.

This technique requires ECG synchronisation, which is used for either prospective or (more commonly) retrospective gating. In retrospective gating (figure 2.10) data are acquired throughout the whole cardiac cycle. During image reconstruction, a retrospective average of the heart rate is determined. Then, data points from longer and shorter RR-intervals are interpolated onto the average of the RR-interval. The acquired data are thus mapped onto a predetermined number of cardiac phases. After reconstruction, the

dynamic images can be displayed in a movie. This method can be questionable in the presence of long beats and beat variations in RR-intervals [24].

Most CINE-MRI images are acquired using a balanced steady-state free precession (bSSFP) imaging sequence, which has an excellent signal to noise ratio and high contrast between blood (bright) and myocardium (dark) [72].

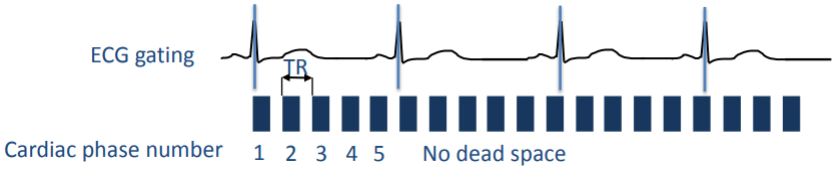

Figure 2.10: Retrospective gating. In this approach, data are acquired throughout the whole cardiac cycle, presenting no dead spaces (when inputs are not collected). The ECG allows the estimation of the heart rate mean. Then, data points from longer and shorter RR-intervals are interpolated onto the average of the RR-interval. TR is the time interval between each acquisition. Image from [24].

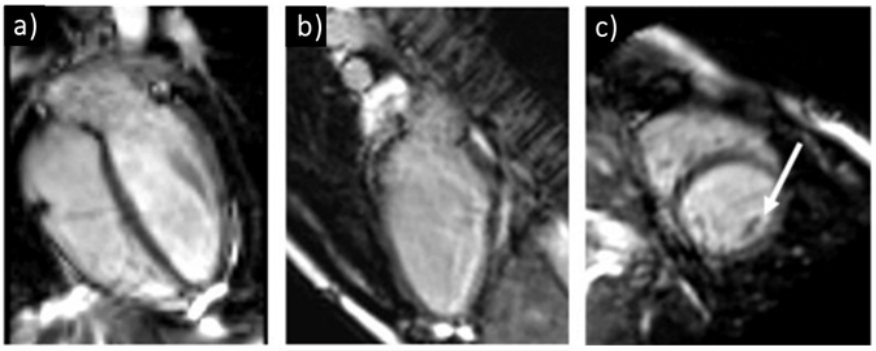

Figure 2.11: CINE-MRI images of a volunteer in diastole showing the **a)** 2-chamber **b)** 4-chamber **c)** short-axis (SA) view of the ventricles (where the arrow indicates the papillary muscles). Image adapted from [25].

2.3.2 DELAYED-ENHANCEMENT CARDIAC MAGNETIC RESONANCE (DE-CMR)

DE-CMR is a highly valuable technique for the diagnosis of MI, to evaluate potential myocardium reversibility and revascularization viability in coronary artery disease, and to distinguish ischemic from non-ischemic myocardial disorders [73]. The underlying concept of this approach is based on the delayed washout of the gadolinium-based contrast agent from injured myocardium areas when compared to the fast washout in normal myocardium [74]. Regarding the typical procedure, a gadolinium-based contrast agent is intravenously injected, and after 10-15 minutes, DE-CMR images are acquired. Typical DE-CMR images are shown in figure 2.12, in the three orthogonal plans. The contrast agent is naturally removed from the body by the kidneys. In regions with scar tissue, the contrast agent washes out more slowly making the areas look bright in T_1-weighted images. In particular, in LV myocardium the late gadolinium enhancement (LGE) can give crucial information about the lesions (scar, infarct, etc) and therefore the risk of an adverse cardiac event and likely response to therapeutic approaches, such as revascularization [13, 75–77]. Furthermore, microvascular obstructions or no-reflow regions (dark areas in T_1-weighted images) have been connected with poorer clinical outcomes.

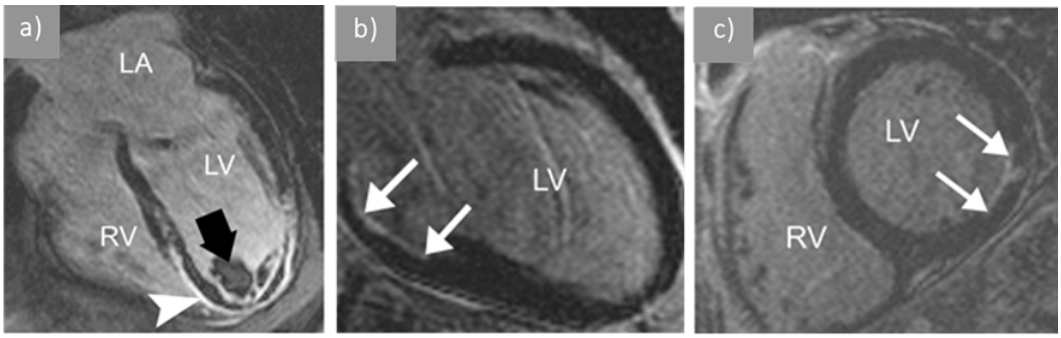

Figure 2.12: DE-CMR images acquired in all three orthogonal planes. LA = left atrium, LV = left ventricle, RV = right ventricle. **a)** 4-chamber view of an acute (48-hour) myocardial infarction (MI) with areas of microvascular obstruction (black arrowhead) surrounded by bright signals (white arrowhead) **b)** 2-chamber and **c)** short-axis (SA) views of an old MI limited by the subendocardial region of the basal inferolateral wall (arrows). Image adapted from [26].

2.4 MANUAL AND SEMI-AUTOMATIC SEGMENTATION

Image segmentation aims to extract the outlines of different regions in an image by dividing it into regions with pixels that have something in common, such as belonging to the same object or tissue. In CMR, segmentation allows the extraction of heart structures such as the LV, using them to quantify metrics as volumes essential to measure vital functions for diagnosis, prognosis and therapy planning. Image segmentation can be performed by several methods:

- **Manual Segmentation:** Manual segmentation refers to the process where the user segments and labels an image file by hand painting. This approach can be very accurate. However, it can be time-consuming and have low reproducibility depending on the region of interest and the segmentation tools at one's disposal.

- **Thresholding:** Thresholding is the process that segments a grey-level image into several distinct regions with different image intensities. The first step creates a histogram by plotting the number of pixels which have a specific grey value versus that value. When normalised, the histogram is essentially the probability density function of the grey values of the image. For example, for a simple binary segmentation task, the user can select a threshold of t_0, label specific pixels (with grey values greater than t_0) as object pixels and the remain pixels (with grey values smaller than t_0) as background pixels.

- **Edge Detection:** This technique examines the differences between neighbouring pixels. The values belonging to distinct regions are then attributed, creating edges that separate them. To measure the dissimilarity between neighbouring pixels, a window moves through the image, calculating the statistical properties of pixels (e.g. image intensity) in distinct halves of the window. After comparing the results, boundaries are identified as places where these statistical properties differ most. For example, in horizontal windows, the local maximum where the value of the difference is greater than a certain threshold, are accepted as possible vertical boundaries between adjacent regions. Then, when rotating the window by 90 and sliding it vertically to scan the whole image again, it will have prior knowledge of the existence of a boundary, which increases the probability of locating the boundaries accurately. Despite really accurate, this method is time-consuming due to the required complex computation.

- **Region growing:** Region growing algorithms, comprising the watershed segmentation method and the split and merge algorithm, take into consideration spatial proximity of pixels. The watershed

method starts from user defined seed pixels and attaches neighbouring pixels to them until all have been assigned to a region (identified by morphological image reconstruction) [78]. This method is a local method, with no global view of the problem, and a continuous path of points related to colour may exist.

- **Atlas-based Segmentation:** The atlas-based segmentation extracts prior knowledge from a reference image, known as atlas, to help the segmentation task. An atlas is defined as the combination of an intensity image (template) and its segmented image (the atlas labels). This method requires the use of image registration (an optimisation problem), in order to align the atlas to an unseen image. Thus, aiming to estimate the transformation between images spaces, this iterative process only stops when obtains a (local) optimum.

  After registering the atlas template and the target image, the easiest and fastest way to assign a label to each input image voxel is to propagate the atlas labels to image space (label propagation). However, this can lead to some errors since the atlas and the subjects are not anatomically equal. To reduce registration errors single atlas can be extended (Multi-atlas label propagation). Thus, when using an atlas database, the outliers can be minimised by discarding those voxels with a low agreement between different label propagation. Finally, in probabilistic atlas-based segmentation probabilistic atlases are used [79].

2.5 EMPIRICAL CLASSIFICATION

Empirical methods to predict the outcome of CAs for AF employ scoring systems. In general, these approaches study different combinations of (discrete or continuous) clinical variables that can be related to AF recurrence, such as a diagnosis of heart failure or the older age. Thus, clinical variables detected in each subject are scored, and the probability of AF recurrence after CA is then calculated by score addition. As long as the estimated score value increases, the probability of AF recurrence is bigger. Consequently, the chance of CA success is lower, and its non-usage can be considered. Some approaches are described below:

- **The ALARMEc Score :** The ALARMEc score was the first reported scoring method aiming to predict the outcome of AF followed CA. The analysis includes score values ranging from 0 to 5 point and five variables, each one scoring 1 point : AF type; left atrial size [normalized left atrial area $\geq$ 10.25]; Renal insufficiency (eGRF $<$ 68 ml/min); Metabolic syndrome; Cardiomyopathy.

- **The MB-LATER Score** The MB-LATER score is proposed to predict very late ($>$12 months) recurrence of AF, after CA, with a range of score values between 0 and 6 points . The five clinical variable used are: Male sex; Bundle branch block; Left atrial size $\geq$ 47 mm; Type of AF (paroxysmal, persistent or long standing persistent); Early recurrent AF.

- **Left Atrial Geometry and Outcome (LAGO) :** LAGO is a combined clinical-imaging score that comprises five items, scoring 1 point each : AF phenotype; Presence of structural heart disease; CHA2DS2-VASc $\leq$1; Dilated LA; Spherical LA. The patients are classified with low ($\leq$2 points) and high risk ($\geq$3 points) of procedural failure. Posterior adjustment and analysis revealed that LA sphericity and AF phenotype are the only strongest predictors of CA outcome [61].

Scoring methods are limited to binarized metrics. Thus, interactions between parameters, such as the influence AF in LA enlargement, and subtle changes are not contemplated. ML scores are more flexible

and therefore expected to be more clinically useful.

2.6 MACHINE AND DEEP LEARNING

ML is considered a sub-field of Artificial intelligence (AI) able to train statistical models and make predictions. Inside ML, DL is a sub-discipline applied mostly to images which can learn directly from data the most useful features for, for example, prediction, segmentation and location tasks.

ML learns firstly which parts of the data are important for a task and then, how to optimally join the information to produce the desired outcome. Learning is the capability to perform a task, for instance, if we want a robot that walks, walking is the task, and we can program the robot to learn to walk or create an algorithm that specifies each step manually. In particular, ML uses supervised and unsupervised learning to train a model for a specific job. In supervised learning, algorithms are trained using datasets where each sample is associated with a label/target, and the aim is to establish specific relationships between the input data and the respective labels. In unsupervised learning, the training procedure employs datasets with unlabeled information, allowing the algorithm to look for relevant patterns without guidance [29].

ML models are defined as mathematical functions with a different capacity depending on the complexity of the tasks. The models' capacity is provided through the degree of the polynomial most suitable for training data. In particular, the ML algorithms can be adopted in different types of tasks, such as regression tasks, where continuous variables are estimated, or classification tasks that predict discrete labels like the presence or absence of diseases [29, 42].

2.6.1 FUNDAMENTALS OF DEEP LEARNING (DL)

DL is a field of ML based on artificial NNs that structures algorithms in layers to progressively extract higher-level features from data. Advances in DL have enabled the development of NNs as the commonly used convolutional neural networks (CNNs) and fully connected networks (FCNs).

A NN is a collection of neurons (figure 2.13) aggregated into layers. The neuronal model returns an output by applying an activation function in a group of weighted inputs. Thus, the key components of a neuron are: an input vector which can be both features from the training set or outputs from a prior layer; a group of synaptic weights; a summing junction with an activation function; and an output. Synapses take input signals, which travel while weights are applied. Then, the neuron sums all the weighted inputs (summing junction) and applies bias (additional constants) that increases or decreases the input of activation functions. Finally, the activation function is employed in the weighted inputs group, and the result is passed to the next layer of neurons [27]. The described operations are mathematically represented as:

$$v_k = W_k X + b_k \tag{2.3}$$

where X are the neurons inputs, k are the activation potentials, W_k are the synaptic weights, b_k is the bias and v_k is the activation potential. Thus, the neuron output is written as y_k:

$$y_k = F(v_k) \tag{2.4}$$

where $F(v_k)$ represents the activation function applied in the activation potential k [27]. In detail, activation functions are mathematical expressions that enable non-linear problems, consequently more

complex, to be solved. They decide if a neuron should be activated or not, and introduce non-linearity into the output of a neuron which is essential for the NN to learn more than linear models. Rectified linear unit activation function (ReLU) is a commonly used function that returns 0 if it receives any negative input, but for any positive value it returns the value back [80], so it can be written as:

$$F(v_k) = \begin{cases} v_k, & \text{if } v_k \geq 0 \\ 0, & \text{if } v_k < 0 \end{cases} \tag{2.5}$$

Parametric Rectified Linear Unit (PReLU) is a generalization of ReLU that allows a network to learn from negative values [81], and can be written as:

$$F(v_k) = \begin{cases} v_k, & \text{if } v_k \geq 0 \\ a_k v_k, & \text{if } v_k < 0 \end{cases} \tag{2.6}$$

where a is a coefficient that controls the slope of the negative part, learned along with the other NN parameters during the training process.

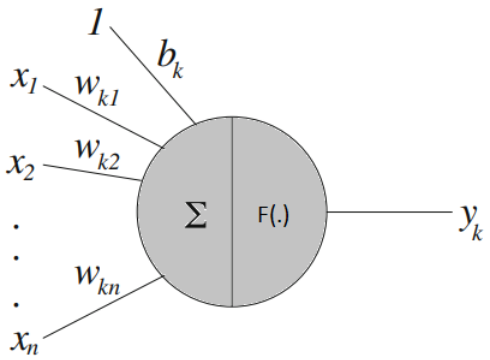

Figure 2.13: Single neuron k and a simple nonlinear model. The inputs are presented by the vector $x = (x1, x2, ..., xn)^T$; b_k denotes the bias; the synaptic weights are represented by $W_k = w_{kj}, j = 1, 2, ..., n$ where j is the index of the n inputs. $F(.)$ is the activation function applied in the group of weights inputs; y_k is the output of neuron k. Image addapted from [27].

Training a network is required for a model to perform inferences, and aims to find the best values of the network parameters. The first step is to initialize the parameters and define the hyperparameters. In particular, hyperparameters are external variables whose value controls the learning process, like the learning rate that guides how much to change the model in response to the estimated error in each iteration. Parameters are model internal variables, like weights and bias from neurons, whose value can be estimated from data. After initialization, starts a loop through a defined number of iterations. In detail, each cycle of the loop takes the following tasks:

1. Forward propagation: Neurons outputs are computed as in eqs. 2.3 and 2.4 to find the NN final output.

2. Loss function estimation: The loss function provides in each training iteration, the network's error regarding predictions (discrepancy between the ground truth and the estimated output).

3. Backward propagation: Method where the gradient of the loss function with respect to the parameters is calculated.

4. Update parameters: Optimizer algorithms change the NNs hyperparameters by analysing the

gradients. In particular, Adam is an optimizer that estimates individual adaptive learning rates for different NN parameters. In detail, this method uses estimations of the gradient first and second moments to adjust the learning rate for each NN weight. The N-th moment of a random variable is the variable expected value to the power of n. This method is straightforward to implement and computationally efficient at the same time [82].

The number of iterations is defined as the number of passes a model goes through a dataset, where one pass comprises a forward and a backward step. Moreover, epochs represent the number of times a network sees the complete dataset. During the training process, the distribution of inputs in every layer is adjusted as the parameters of prior layers variates, making the NN training a complex task. Thus, normalization approaches have been developed to establish an equal scale when features present different ranges. In particular, instance-Layer normalization, detailed in [83], uses the sigmoid function for the feature map combination and cascading group normalization.

Commonly used loss functions for image processing tasks include the Mean Squared Error (MSE), the cross-entropy (CE) and the Dice loss function [28]. MSE is the simplest loss function normally used in regression tasks such as heart localization, calcium scoring, landmark detection and image reconstruction:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{2.7}$$

where y_i is the target vector values, $\widehat{y}_i$ is the vector of the predicted values and n is the number of data samples. The CE loss function is given by the following expression:

$$L_{CE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_i^c log(p_i^c), \tag{2.8}$$

and the Dice loss function is described as follows:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{n}\sum_{c=1}^{C} y_i^c p_i^c}{\sum_{i=1}^{n}\sum_{c=1}^{C}(y_i^c + p_i^c)}, \tag{2.9}$$

where C is the number of all classes, p is the predicted probabilistic output and y the corresponding target map for each class c.

To develop a DL model large datasets are required, containing paired images and labels $\{x, y\}$, which are divided into training, test and validation sets. The training set is a dataset with samples used during the learning process to choose/optimise the parameters. Moreover, the validation set is used to select the model that has the best ability to generalize over unseen data. After the training procedure, the models can be used to analyse data they have not seen (test set). The parameters values control the behaviour of the system, and a successful model presents a high performance in the job even in the presence of new data.

A conventional CNN is shown in figure 2.14. It includes an input layer, hidden functional layers to transform the data into a specific form like vectors, and an output layer.

The functional layers usually comprise pooling layers, convolutional layers and/or fully-connected layers. In particular, the convolution operation is denoted as:

$$s(t) = (x * w)(t) \tag{2.10}$$

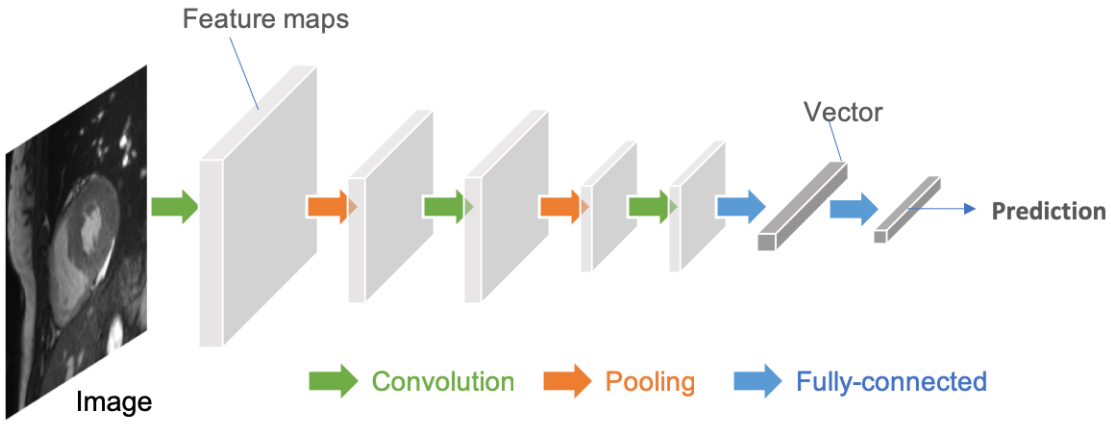

Figure 2.14: Representative structure of a CNN used in a classification task. A generic CNN learns hierarchical features by employing a stack of convolutions and pooling operations using, in this example, cardiac MR images as inputs. Then, fully connected (fc) layers reduce the spatial feature maps into a flattened vector. Image from [28].

where the first argument x is the input and the second argument w the kernel. In figure 2.15 a 2-D convolution is represented. For 2D and 3D inputs, kernels of size n x n and n x n x n are used, respectively. Convolutions are applied to extract features, and then batch normalization is used to obtain feature maps by passing the output through a nonlinear activation function. Then, pooling layers aggregate neighbour pixels, which reduces the feature maps dimensions, decreases the number of parameters to learn and the amount of computation performed in the network. Standard approaches comprise the average pooling, that calculates the average value for each patch on the feature map, and the maximum pooling, which estimates the maximum value. The final sized feature maps are the input of the next convolutional layer. In classification tasks, fully connected layers are usually used to compute a score for all feature maps and a 1D array as output. The output presents a size equal to the number of classes, and the softmax function converts the score into a class probability. In detail, softmax applies the exponential function to each element of input vectors normalizing the values by dividing them to the sum of all exponentials. After softmax, each component belongs to the interval (0,1). Softmax function is mathematically defined as follows:

$$\sigma(\vec{z_i}) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} for\ i = 1, ..., C\ and\ z = (z_1, ..., z_C) \in \mathbb{R}^C \tag{2.11}$$

where $\vec{z_i}$ is the input vector and C represents number of classes in the multi-class classifier [29].

In some tasks, CNNs have an encoder-decoder composition producing an output with the same input size. Firstly, the encoder modifies the input image into a low-dimensional feature representation. Then, the decoder interprets the feature maps and applies transposed convolution and convolution operations to recover the image space details for pixel-wise prediction.

Many variants of CNNs, have been recently developed, in particular, U-net (figure 2.16) presents a contraction, a bottleneck, and an expansion section, which converts an image into a vector and uses the respective mapping to turn it again into an image. In the example shown in figure 2.16, the contraction part comprises four contraction blocks, each one applies two 3 x 3 convolutions ReLu layers (blue rows in figure 2.16) to the input and then a 2 x 2 max pooling layer (red rows in figure 2.16) which reduces the number of feature maps. The bottleneck section employs two 3 x 3 ReLu convolutions layers and a 2 x 2 upsampling convolution (green rows in figure 2.16). Finally, the expansion layer splits the number

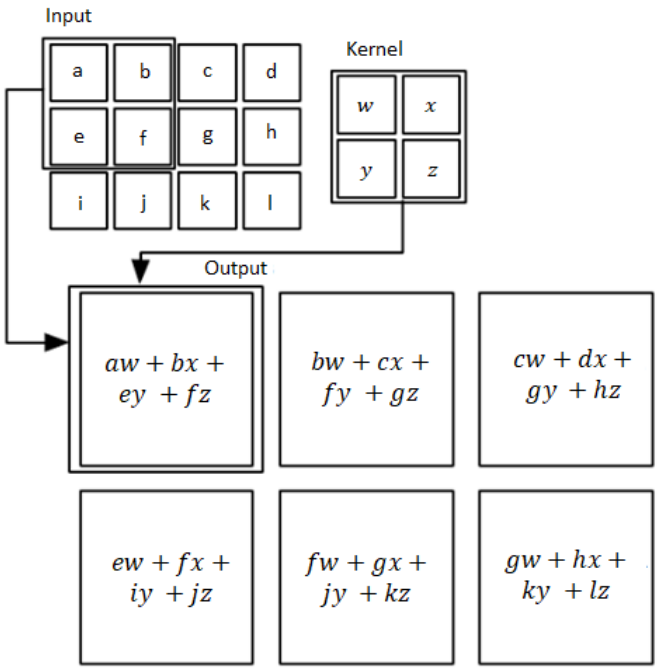

Figure 2.15: 2D convolution. In this example, the upper-left element of the output tensor is formed by applying a 2 x 2 kernel to the corresponding upper-left region of the input tensor. Image from [29].

.

of feature channels by moving the input into blocks with two 3 x 3 convolutions layers and a 2 x 2 upsampling layer. The final step includes a 1 x 1 convolution layer (light blue rows in figure 2.16) to make the number of feature maps equal to the number of segments desired in the output [30]. The U-net structure is commonly used in the field of image segmentation. In a classification task, the softmax function is applied for each image pixel followed by a loss function, which helps identify individual cells within the segmentation map, and converts automatic segmentation into a classification problem where the task is to classify each pixel to one of the classes.

Adding multiple layers can saturate the network's performance leading a vanishing gradient problem. Thus, residual Networks (Res-Nets) use residual units or blocks, giving an alternative path for the gradient to flow through. In detail, a residual unit presents skip connections that link nodes from different layers bypassing one or more layers of nonlinear processing. Figure 2.17 shows a residual block example with two layers followed by a ReLU activation function. The skip connection is a shortcut, jumping both layers with an identity function and performing a element-wise addition to ReLU [31].

Generalization is the models' capability to work well on new data and it is quantified using the generalization error. In detail, the generalization error is estimated by measuring the error performance on the validation set (green line on figure 2.18), which is always higher than the training error (estimated in the training set and represented in blue on figure 2.18). The final goal is not only to find the lowest training error but also the smallest gap between both values, which is named the optimal capacity (red line in figure 2.18). When not in optimal capacity, two scenarios can happen, underfitting and overfitting. In underfitting conditions, the training error is not the lowest while overfitting happens if the gap between the training and validation errors is too large. The example of figure 2.19 shows that in underfitting scenarios, algorithms often have a low capacity and cannot efficiently capture the underlying data structure. On the other hand, in an overfitting scenario, where the models' capacity is too high, algorithms fit extremely well on training data but are unable to perform well on test sets.

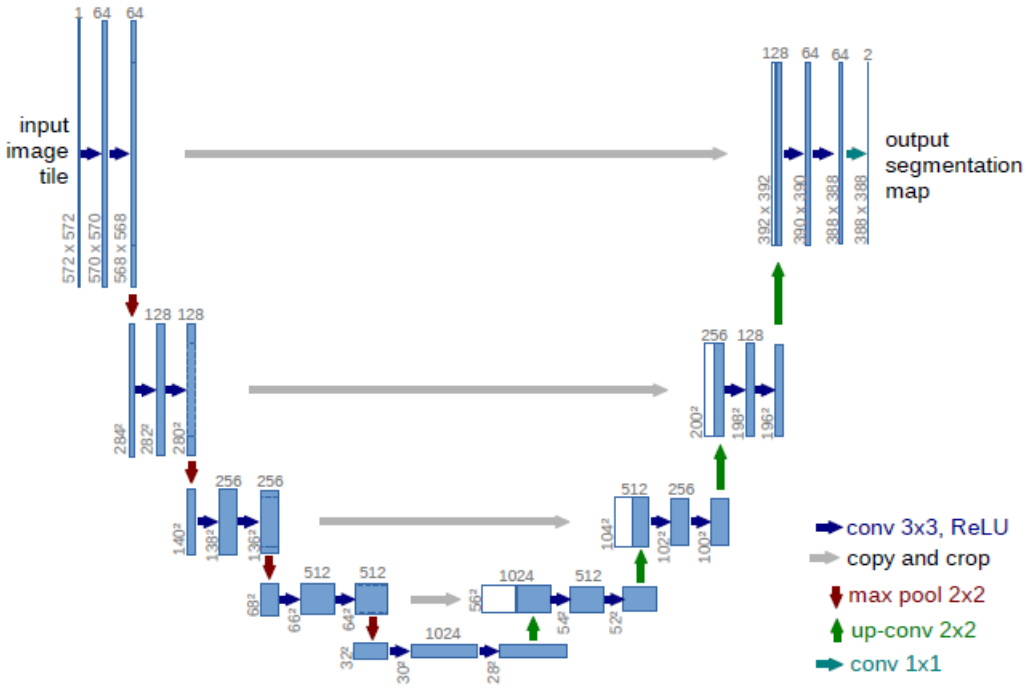

Figure 2.16: U-net architecture with 32x32 pixels in the lowest resolution. While blue boxes correspond to a multi-channel feature map, the white ones represent copied feature maps. On top of the box is indicated the number of channels whereas in the lower left edge is provided the x-y size. The blue, grey, red, green and light blue arrows respectively represent 3 x 3 convolutions ReLu layers, copy and crop operations, max pooling layers, 2 x 2 upsampling convolutions and 1 x 1 convolution layers. Image from [30].

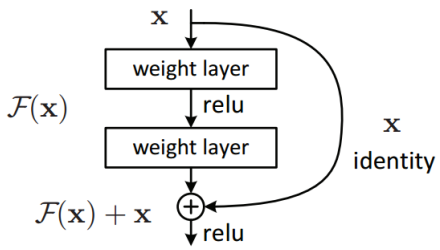

Figure 2.17: Single residual block representation. x is the input vector and $F(x)$ represents the residual mapping to be learned. The operation $F + x$ is performed by a shortcut connection and element-wise addition. Image from [31].

Achieving high performances is a challenging task, so different optimization and regularization approaches can be applied. While optimization focuses on minimizing the training error, regularization methods, such as dropout, try to reduce the generalization error. In detail, dropout (Figure 2.20) works in a probabilistic manner, where p is the probability of a neuron being ignored at each layer during the training phase. This technique reduces overfitting since neurons from the same layer can share data and overload the final parameters' values, thus depending on one another [29].

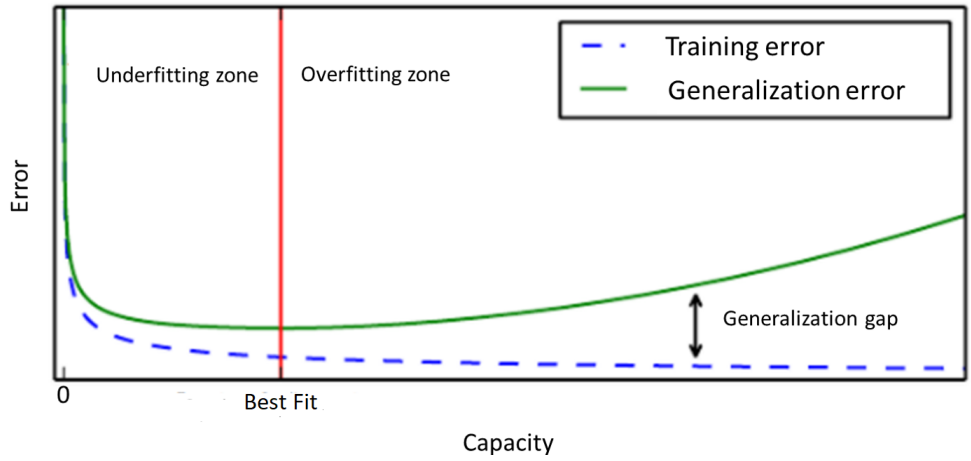

Figure 2.18: Graphic illustration of how generalization and training errors, represented by the green and dashed blue lines, respectively, depend on a network's capacity. The capacity of a model describes its complexity. The red line represents the optimal capacity, where the model best fits data. On the left end of the image, the underfitting domain is displayed where both error values are high. On the right side, the overfitting scenario shows an increase in the gap between errors. Adapted from [29].

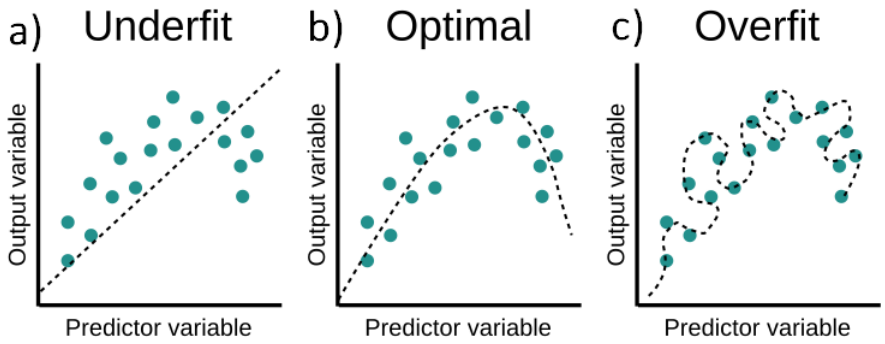

Figure 2.19: Graphic representation of underfitting, overfitting and appropriate fitting. **a)** Underfitting: the model may not capture accurately the underlying structure of the data; **b)** Optimal capacity: denotes good fitting ; **c)** Overfitting: the model function fits too closely the training data that fails on predicting accurately additional data. Image adapted from [32].

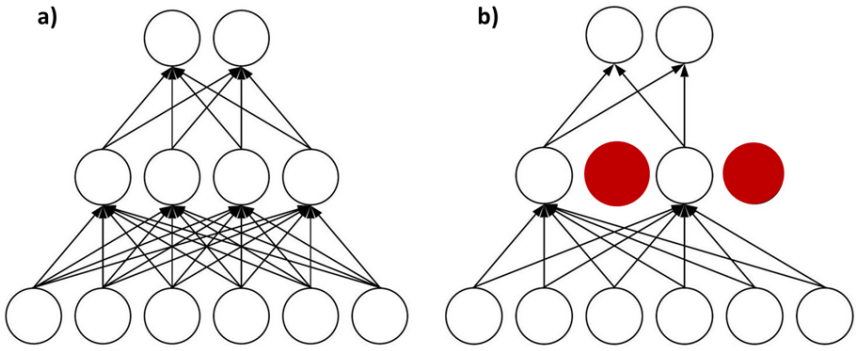

Figure 2.20: **a)** Simple NN with no dropout (all neurons work and share weights among them). **b)** NN applying a regularization method, dropout. Red neurons are ignored, creating fewer connections between neurons. In subsequent training iterations, other neuronal connections will be ignored (keeping the fraction of removed connections constant) to discourage networks with a high capacity from overfitting to training data. Image adapted from [33].

2.6.2 QUANTITATIVE METRICS

To quantitatively assess the performance of automated segmentation models, surface distance-based metrics as the median contour distance (MCD), the Hausdorff distance (HD) and the Dice coefficient (DC) are commonly used. The DC estimates the overlap ratio between a automated segmentation (A) and a manual segmentation (B), ranging from 0 (mismatch) to 1 (perfect match):

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{2.12}$$

The HD measures the maximum distance separating the segmentation contours δA and δB:

$$HD = max\Big(\underset{p \in \delta A}{max}\, d(p, \delta B), \underset{q \in \delta B}{max}\, d(q, \delta A)\Big) \tag{2.13}$$

Finally, the mean contour distance computes the mean distance also respectively to δA and δB:

$$MCD = \frac{1}{2|\delta A|} \sum_{p \in \delta A} d(p, \delta B) + \frac{1}{2|\delta B|} \sum_{q \in \delta B} d(q, \delta A) \tag{2.14}$$

where $d(p, \delta)$ indicates the minimal length from point p to contour δ. It is important to stress that for HD and MCD, the lower the distance metric, the better the agreement [42].

Accuracy, sensitivity, and specificity metrics can be used to evaluate models in the classification task. In particular, sensitivity is the ability to correctly identify true positives (TP), whereas specificity is the capability to classify true negatives (TN). Moreover, accuracy is defined as the fraction of predictions the model got right [84]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.15}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.16}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.17}$$

where TP, TN, FP, FN are respectively the true positives, true negatives, false positives and false negatives classification cases. The confusion matrix shown in table 2.2 is also a representation used to assess classification algorithms.

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Table 2.2: Confusion matrix is a table with two rows and two columns that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN).

2.6.3 STATE-OF-THE-ART DEEP LEARNING (DL) METHODS FOR CMR SEGMENTATION AND CLASSIFICATION

In the image segmentation field, novel methodologies have been developed during the past few years. In particular, 2D networks using a slice-by-slice manner to segment the atrium from 3D LGE images were developed by Xiong et al. [34], Preetha et al. [85] and Bian et al. [86]. Xiong et al. [34] created AtriaNet, the dual FCN represented in figure 2.21, to automatically segment the LA epicardium and endocardium. The dual pathway obtained not only the local atrial tissue geometry, but also the global positional information. The architecture employed thirteen successive convolutions and three further convolutions to merge data. The final output presented two feature maps with the indication of positive/negative pixel classification probability. Preetha et al. [85] followed a U-Net structure, using skip connections to provide high-resolution features to the decoding path. The method also included deep supervision operations, which used single task labels to control hidden layers [87] and thus relieve gradient vanishing and information losses through forward propagation. Moreover, Bian et al. [86] presented a pyramid network with online hard negative example mining loss to help the network in samples selection. The method used 2D slices as inputs and consecutive residual blocks to extract hierarchical features. Then, the pyramid module collected semantic cues in feature maps from multiple scales for fine-grained segmentation.

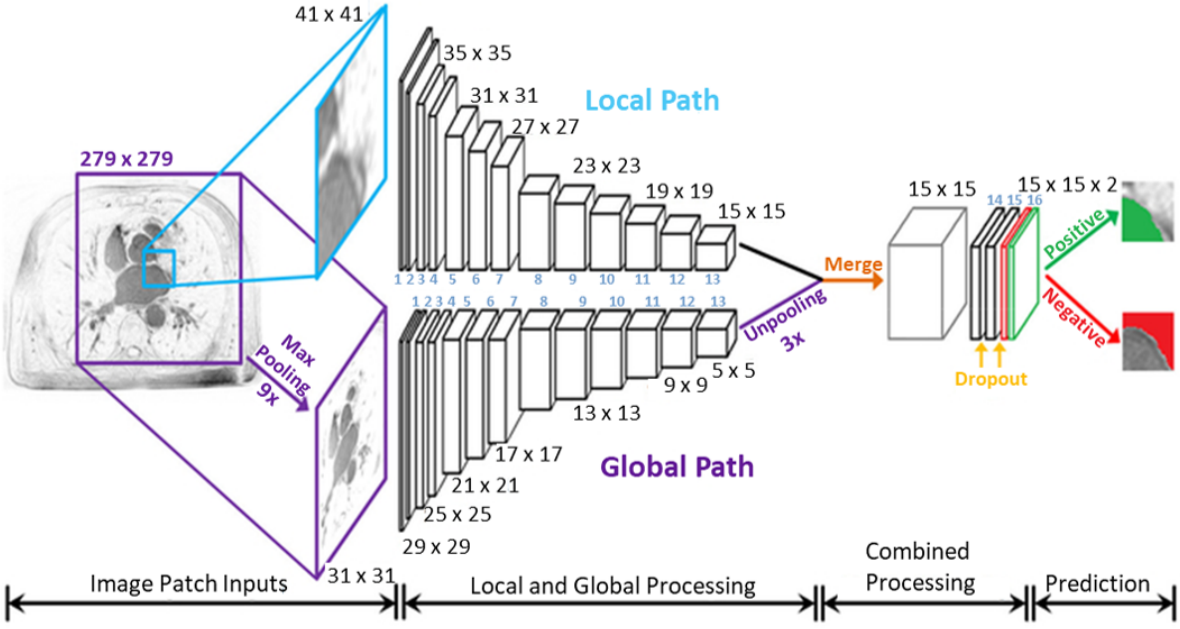

Figure 2.21: AtriaNet architecture. The network presented two FCN for global and local pathways. Three further convolutions were applied to merge data in the combined processing. The final output had two feature maps for every 15 x 15 patch, presenting the probability of positive/negative pixel classification. Image from [34].

3D approaches for atrium segmentation from 3D LGE images were developed by Savioli et al. [88] and Xia et al. [89] using volumetric fully convolution neural networks (V-FCNNs). A V-FCNN presents two main paths, the volumetric down-sampling path and the volumetric up-sampling path. Savioli et al. [88] applied four 3-D convolutions blocks followed by PreLU along with 3-D batch normalisation layers in the volumetric down-sampling path to progressively reduce the size of each input slice. Then, 3-D up-sampling convolution blocks followed by PreLU and 3-D batch normalisation restored the compressed volume to its initial size. Xia et al. [89] created two V-FCNNs for 3D automatic atrial segmentation. The

first one was used to find a fixed region that covered the entire atria. The second approach segmented the obtained cropped target region and transformed the predicted masks in the target region to the original size volume.

Jia et al. [90], Vesal et al. [91] and Li et al. [92] also created 3D methods for atrium segmentation from 3D LGE images by applying U-Nets. In particular, Jia et al. [90] used successive 3D U-Nets and a contour loss. The first U-Net coarsely segmented and located the LA while the second U-Net precisely segmented the structure under higher resolution. Vesal et al. [91] added dilated convolutions in a 3D U-Net which established the space between values in a kernel. Li et al. [92] created a 3D U-Net network named attention based hierarchical aggregation network (HAANet). An attention mechanism was adopted to improve the efficiency of extracting features, while hierarchical aggregation provided the capability of shallow and deep feature fusion. The network was composed of an encoder and a decoder path. The hierarchical aggregation unit and the attention unit composed attention based hierarchical aggregation modules (HAAMs) applied as stacked series in the encoder path. The decoder path followed the usual U-Net structure. The method used each stage of the two convolutional layers as a region of interest (ROI) detection network to segment the area containing the whole 3D LA volume. After being cropped out the ROIs were fed in the HAANet for the final LA segmentation.

Finally, Mortazi et al. [93] and Yang et al. [94] segmented the atrium from 3D LGE images using multi-view FCNs. All the presented NNs segment the atrium using LGE MRI images, which compared with CINE MRI have a distinct contrast and resolution. In particular, CINE MRI combines good spatial and temporal resolution with large cardiac coverage and is thus ideally suited for assessing LA function. Thus, in this work, a dedicated NN is proposed for fully automatic SEGmentation of the left Atrium in short axis CINE MRI based on a CNN (SEGANet), detailed on section 3.

In the classification domain, DL's first steps were done to predict two classes, usually healthy and non-healthy cases. In the early days, the lack of available data stimulated the usage of transfer learning techniques, which apply pre-trained networks to perform similar tasks. In particular, fine-tuning is a common technique for transfer learning that uses weights of a previous DL network to program another similar process. Good results were firstly achieved by Esteva et al. [95] and Gulshan et al. [96] when fine-tuning was used in a pre-trained version of GoogleNet Inception v3 [97] to respectively classify skin cancer lesions and detect retinopathies in retina images.

Nowadays, DL methods are widely used for medical image analysis. DL models can learn patterns of health trajectories from vast numbers of patients, which allows the possibility to use them in the diagnosis and prognosis field, anticipating future events. In particular, a DL approach to identify cases with good prognosis of cardiac resynchronization therapy was developed by Peressutti et al. [98]. Bello et al. [35] also adopted DL in CINE-MRI to predict human survival due to heart disease, and Chen et al. [99] proposed a CNN that performed pre/post ablation classification.

In detail, Peressutti et al. [98] presented a framework that combined a cardiac motion atlas with non-motion data. The atlas described the cardiac cycle motion from a population sample, and was created with the following steps: Estimation of LV geometry and motion; Spatial normalisation of LV geometries; Motion deformation; Reorientation to the atlas coordinate system; and projection of the high dimensional matrices into random sparse matrices. The non-motion data were obtained from different sources such as imaging, ECG and clinical reports. Afterwards a supervised learning approach with a multiple kernel learning (MKL) algorithm was trained to learn the relationship between the atlas data and known clinical outputs. The motion of the heart was also tracked by Bello et al. [35] using an automatic segmentation of the ventricles. In detail, the architecture of the segmentation algorithm comprised a FCN, predicting the

segmentation and producing location probability maps. The results were then applied to track the motion of corresponding anatomic points, forming a dense motion model of cardiac motion. The CMR-derived features were then joined with clinical information to train a DL classifier and predict human survival due to heart disease. The prediction algorithm, a supervised denoising autoencoder (figure 2.22), used the resulted dense motion model as input and applied dropout as a regulatization approach. Then, data were sequentially fed in hidden layers, being compressed and upsampled for feature extraction.

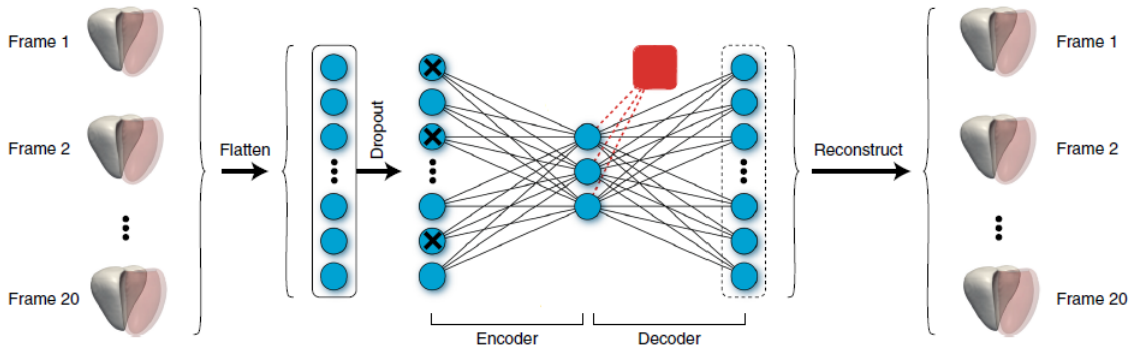

Figure 2.22: Overview of the method proposed by Bello et al. [35]. The autoencoder takes cardiac motion meshes as its input (RV shown in solid white, LV in red). Then, the data is flattened and dropout is applied, entering in the hidden layers. The red dot represents the central hidden layer. Finally, the data is reconstructed.

Finally, Chen et al. [99] proposed a CNN to perform both atrial segmentation and pre/post ablation classification. While inputs were 2D DE-MRI images, outputs were LA segmentation masks and pre/post ablation classification scores. The segmentation method structure was derived from the 2D U-Net. Classification task used features learned after the 4th max-pooling layer in the down-sampling path of segmentation. Spatial pyramid pooling was applied to fix the length of the vectors, which were then processed in FC layers with a dropout of 0.5, followed by a softmax layer to calculate class probabilities for each image.

The progress in automatic classification has shown the potential of using these approaches as clinical tools for diagnosis and prognosis. However, DL has never been used to perform a complete 3D analysis of LA dynamic characterisations and volumes across the whole cardiac cycle. In particular, LA dimensions and shape have revealed to be reasonable predictors of post-ablation patient outcome [7]. Thus, this project aimed to develop a classifier able to predict AF recurrence after CA in AF patients using pre-ablation LA short axis in CINE MRI images (section 5).

3 SEGANET

The network for Fully Automatic SEGmentation of LA from short axis CINE MRI (SEGANet) was developed as detailed in `https://arxiv.org/abs/2008.13718`. The current chapter introduces the methodology used to achieve the LA segmentation goals, focusing on CINE data acquisition, the SEGANet architecture and the employed training procedures.

3.1 METHODOLOGY

3.1.1 DATA ACQUISITION

Short axis CINE MRI image stacks covering all phases of the cardiac cycle, were obtained using a 1.5T Philips Ingenia scanner with a 32-channel cardiac coil. For this task, a 2D bSSFP protocol (flip angle: $60°$, TE/TR: 1.5/2.9 ms, SENSE factor 2) was adopted. In particular, images were acquired with ECG-based retrospective gating, a field of view of 385 x 310 x 150 mm^3, an acquisition matrix of 172 x 140 and a slice thickness of 10 mm. Then, images were reconstructed to a resolution of 1.25 x 1.25 x 10 mm^3, 30-50 Cardiac Phases and $60 - 70\%$ of view sharing.

The acquisition was performed (under permission and following informed written approval) in 60 AF patients (75% male) between 31-72 years old, and in 12 healthy volunteers (50% female) from 24 to 36 years old. These acquisitions offered complete coverage of the ventricles and atria, comprising 4-6 of atria slices (Figure 3.1). Every slice was collected in a separate breath-hold. In contrast, most CINE scans (e.g. the UK Biobank protocols [100]) typically only image the ventricles. Hence, the CINE acquisitions performed in this research project are quite unique, since they image both the ventricles and atria.

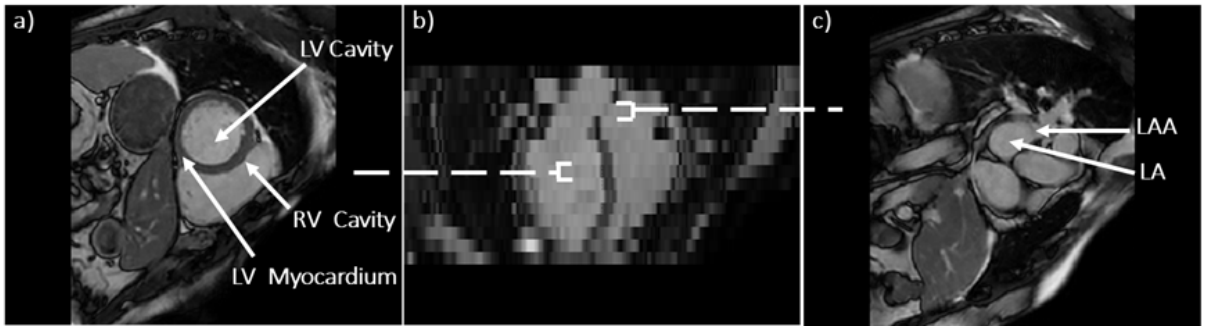

Figure 3.1: CMR Cine slice from the project dataset: **a)** Representative short axis slice of the ventricles with right ventricular (RV) cavity, left ventricular (LV) cavity and LV myocardium indicated by arrows; **b)** Pseudo 4-chamber comprising the atria and ventricles. The location of the short-axis (SA) slices displayed in **a)** and **c)** is also indicated ($\rightarrow$) **c)** Representative short axis slice of the atria with the left atrial (LA) body and the LA appendage (LAA) indicated by arrows.

3.1.2 IMAGE PREPROCESSING

Since the segmentation task was developed with supervised learning, the images had to be labelled so the network could map an input (CINE images) to an output (LA segmentation). To obtain the ground truth (GT or M1), manual LA segmentations were performed by two medical specialists on a slice by slice basis at 3 phases of the cardiac cycle: AS, ES and ED. The LAA was included in the segmentation, but the Pv insertions were not.

All the short axis CINE MR slices were intensity normalized and cropped in plane to a matrix of 256 x 256. The dataset was randomly split into 3 sets: training, validation and test (Table 3.1). The selection was performed separately in each class, and then data were concatenated in a final matrix to ensure the same class balance across all three groups. The distribution was done by randomly splitting the cases, not the image slices, avoiding the presence of information from the same input in different datasets, which could bias the model's performance.

	Training Set	Validation Set	Test Set
Subjects	52	7	13
Slices	715	126	234
Percentages	70%	10%	20%

Table 3.1: Details from the division of the 72 cases accordingly to a 0.70/0.10/0.20 ratio.

The images across the cardiac cycle and the corresponding labels were loaded as 3D matrices in the following configuration :

$$T_I = (B \times H \times W) \tag{3.1}$$

where, B represents the minibatch size, H the image height and W the width. For the segmentation task the following was used: B = 300, H = 256, W= 256.

3.1.3 DATA AUGMENTATION

To regularise the network, prevent overfitting and obtain more generalised information from the relatively small dataset, several randomly-chosen data augmentation functions were applied to each slice. These data augmentations were selected to reproduce realistic variations in images without modifying the cardiac features of interest.

We use the following data augmentation operations:

1) Translations, flips, shifts and rotations of 90^o,180^o and 270^o. In detail, image arrays are rotated 90 degrees in the plane one, two or three times (Figure 3.2 d)). Translations swap x and y axes (Figure 3.2f)). Flips reverse the order of elements in an image array, along a randomly chosen axis (Figure 3.2e)). Arrays are also randomly shifted (up, down, left or right) by fractions of the array dimensions (Figure 3.2i)).

2) Additive stochastic noise (Figure 3.2c)).

3) k-space corruption to simulate acquisition artefacts. The function distorts arrays by applying dropout in the k-space. The dropout factor is calculated with a uniform distribution (Figure 3.2g));

4) Smooth non-rigid deformations encoded using free-form deformations as pairs for smooth distortions; The function deforms arrays randomly with a deformation grid (Figure 3.2h));

5) Intensity scalings with a random factor in rectangular image zones (Figure 3.2b)).

3.1.4 SEGANET STRUCTURE

SEGANet is a NN based on the U-Net architecture [30], which contains four layers stacked together, combining the encode and decode paths, as illustrated in Figure 3.3. The network has 2 residual units with a kernel size of 3. Convolutions with a stride of 2, downsample data in the encoding path, and the output tensors of the layers have channel dimensions of 16, 32, 64, 128, and 256. In detail, the network presents several modifications relative to the U-Net: 1) Through the encode and decode paths SEGANet

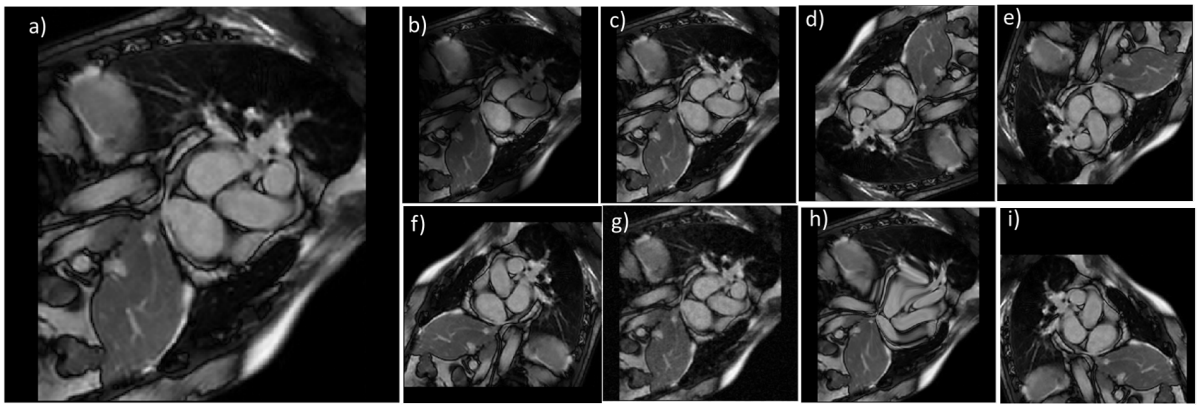

Figure 3.2: Augmentation used during the training process. Representative 2D slice with: **a)** no augmentation applied; **b)** intensity scaling **c)**; additive stochastic noise; **d)** 180^{o} rotation; **e)** 2D vertical flip; **f)**Translation; **g)** K-space corruption; **h)** Smooth non-rigid deformations; **i)** Imaging shift. All the augmentation functions were implemented in slice a).

has residual units [101], which prevent vanishing gradients by enabling the inclusion of measures from the original input in the layer's output; 2) instance layer-normalization, which reduces the influence outlier images may have in the entire batch; 3) Parametric Rectified Linear Unit (PReLU) activation, that permits the network to learn from negative values.

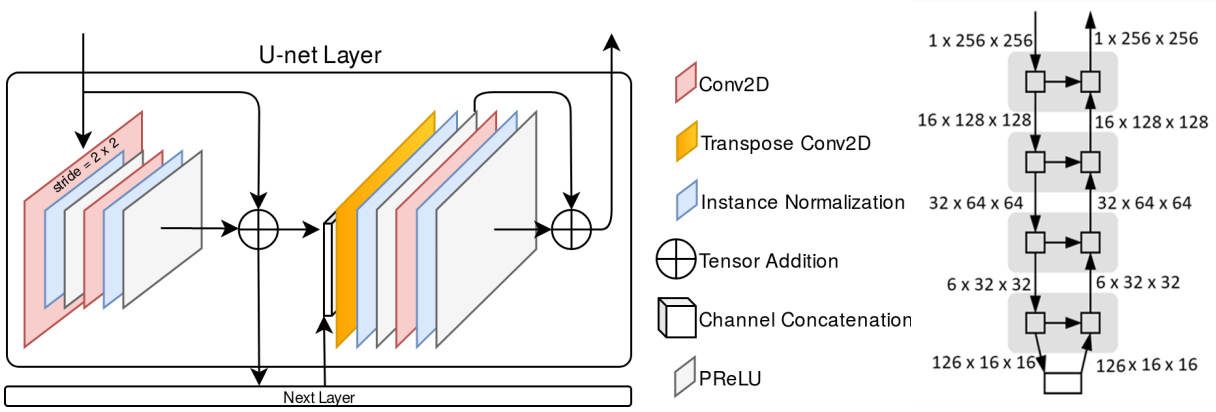

Figure 3.3: The proposed LA segmentation network SEGANet is based on a U-Net architecture. The segmentation network is built as a stack of layers (shown on the right), which combine the encode path of the network on the left and the decode path on the right. Subsequent layers, or the bottom connection, are denoted by "Next Layer". The bottom connection is composed of a single residual unit including two sets of convolution/normalisation/regularisation sequences. Image from [36].

Training Process The network was trained for LA segmentation on short axis CINE MRI throughout $10,000$ train steps and 5 substeps ($50,000$ iterations). During the process mini-batches of 300 slices were drawn randomly from the 715 slices of the training dataset. The applied hyperparameters, detailed in table 3.2, included the binary formulation of the dice loss function [102] and the Adam optimiser [103] with a learning rate of 0.0001. The loaded inputs were 3D tensors, as described in 4.3. The LA area was considered as the only label (Number of Classes = 1). The method was implemented using Python (version 3.7) and PyTorch (version 1.7.0) in a workstation with a Titan Xp 12GB card. During the training process, data augmentation was applied in a randomly selected batch, each function with a 50% chance of being applied in a single image, producing a total of 321,750,000 images.

To evaluate SEGANet performance, additional manual LA segmentations (M2) in 13 subjects (test set) were also provided by an additional medical expert. Then, using the ground truth, segmentation quality metrics (DC, HD and MCD) were calculated relative to the SEGANet segmentations and the

Training Configuration	
Parameter	Value
Learning Rate	0.0001
Train Steps	10000
Substeps	5
Batch Size	300
Image Size	256 x 256 pixels
Number of Classes	1
Channels	(16,32,64,128, 256)
Strides	(2, 2, 2, 2)
Residual Units	2
Kernel Size	3

Table 3.2: Hyperparameters used in SEGANet.

manual segmentations M2. The agreement between the two sets of manual segmentations, GT and M2, was also estimated.

3.1.5 POST-PROCESSING

In datasets with a large slice thickness (such as the 10-mm ones used this analysis), it can be hard, even for an expert, to recognise the separation between the atria and ventricles. Thus, the segmentation obtained from a pre-trained ventricles segmentation network [104] was used to distinguish ventricular tissue from the atrial tissue. The model divided the images slices into three categories (Label 1: RV; Label 2: LV; Label 3: LV Myocardium). By identifying the slices containing the ventricles it was possible to automatically select the ones covering the atria and use them in the analysis of the atrial function (section 3.1.6).

3.1.6 DATA ANALYSIS

To demonstrate the clinical applications of the automatic CINE MRI analysis carried out, LA and LV volumetric time curves were plotted for complete visualisation of the heart volume variations throughout the cardiac cycle. Then, LA EF and aEF values were computed for both AF patients in sinus rhythm and healthy subjects as described in 2.1 and 2.2.

3.2 RESULTS

For the segmentation of the LA, SEGANet was trained with LA images and segmentations following the training configuration presented in Table 3.2.The method achieved 0.077 in the (Dice score) loss value for the training set and 0.083 for the validation set as shown in Figure 3.4.

As already referred, ventricular segmentation in the short axis CINE MRI image stacks allowed the automatic selection of slices covering the atria. In particular, figure 3.5 shows three orthogonal CINE MRI views for one representative patient in the test set overlaid with the automatic LA segmentations after post-processing. The high-quality of the automatic segmentation for the ventricles and atria can also be seen in figure 3.6, where the same CINE MRI is overlaid with all segmentation labels and a 3D volume rendering is displayed. The LA, RV, LV and LV myocardium segmentation are respectively represented in red, blue, yellow and green. Comparing the SEGANet performance with the GT and the additional

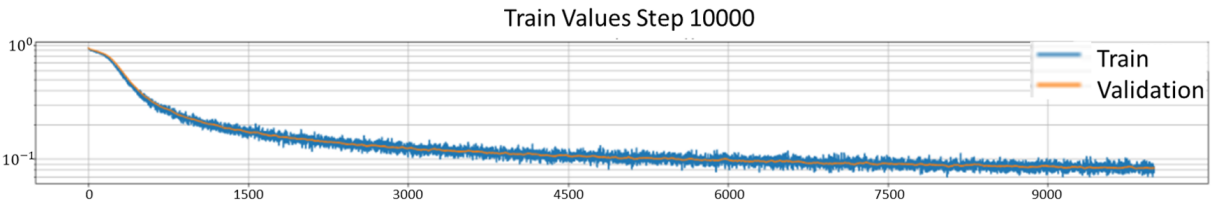

Figure 3.4: Training (blue curve) and validation (orange curve) loss curves for SEGANet, during 10 000 iterations. The training value was calculated at every iteration, while the validation loss was estimated at the end of every epoch.

M2 segmentations, the results, in figure 3.7, show that automatic segmentations are on par with manual segmentations M2, achieving DC: 0.93 ± 0.04, HD: 4.59 ± 2.06 mm; MCD: 0.75 ± 0.31 mm for the SEGANet segmentations when compared to the GT.

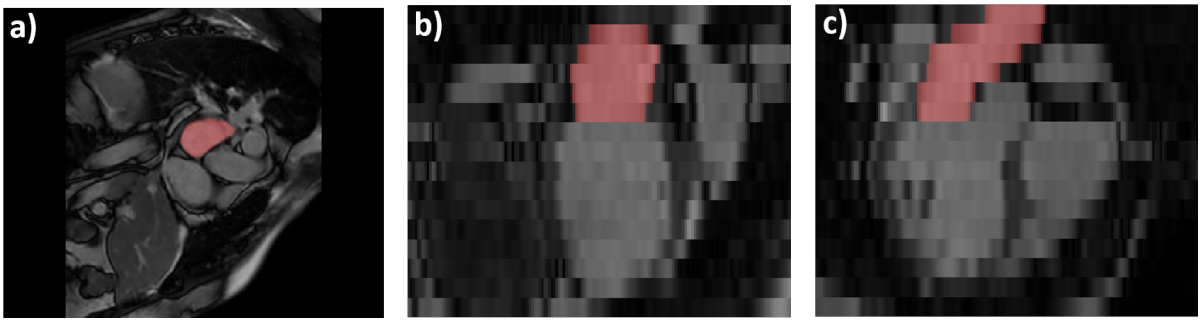

Figure 3.5: LA segmentations obtained with SEGANet for one representative subject during atrial diastole, overlaid on the CINE image in: **a)** short axis, **b)** pseudo 2-chamber, and **c)** pseudo 4-chamber views. The acquired short axis CINE image stacks include both ventricles and atria. Image from [36].

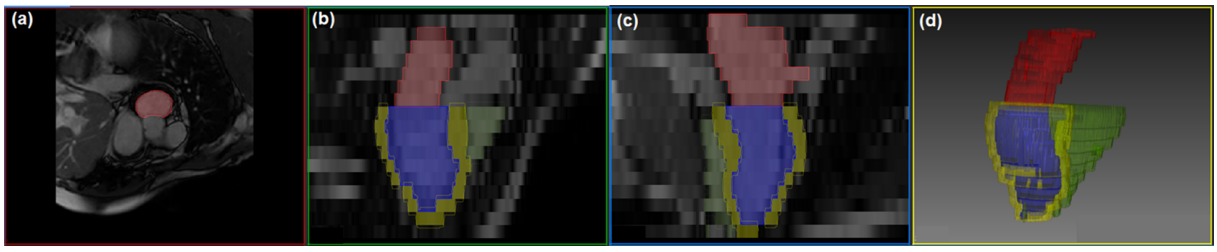

Figure 3.6: **a)** Short axis, **b)** pseudo 4-chamber, **c)** pseudo 2-chamber views and **d)** 3D volume rendering of the left atrium (red), right ventricle (green), left ventricle (blue) and left ventricular myocardium (yellow) automatic segmentations from one representative subject, overlaid on the CINE image. The acquired short axis CINE image stacks include both ventricles and atria.

The LA and LV volumes vary smoothly across the cardiac cycle, further highlighting the good performance of automatic segmentations in all the cardiac phases. In detail, Figure 3.8, shows the LA curve throughout the cardiac cycle for an AF patient and a healthy subject identifying the maximal volume (V_{max}), the minimal volume (V_{min}) and the second peak volume (V_{preA}). Furthermore, figure 3.9 presents both LA and LV curves for two additional representative subjects with the different the cardiac phases labelled.

The estimated LA volumes for AF patients (V_{min} : 79.40 ± 25.34 mL; V_{max} : 111 ± 24.00 mL; V_{preA} : 103.40 ± 25.34 mL) and for healthy subjects (V_{min} : 22.43 ± 25.34 mL; V_{max} : 44.23 ± 24.73 mL; V_{preA} : 35.47 ± 19.48 mL) demonstrate high inter-subject variability, on pair with the literature [105].

The obtained LA EF values for AF patients ($31.1\% \pm 9.9\%$) and healthy volunteers ($49.8\% \pm 7.6\%$) are also in good agreement with published echocardiographic and 2D CINE MRI values [105–107]. aEF estimates (AF patients: $24.3\% \pm 9.0\%$, healthy volunteers: $37.9\% \pm 10.1\%$) are also in agreement with the literature [8, 105, 107]. As expected, the results are significantly higher in healthy subjects than in AF patients for both LA EF and aEF ($p < 1e^{-4}$ and $p < 1e^{-7}$, respectively) as shown in figure 3.8b) and 3.8d) [8, 106].

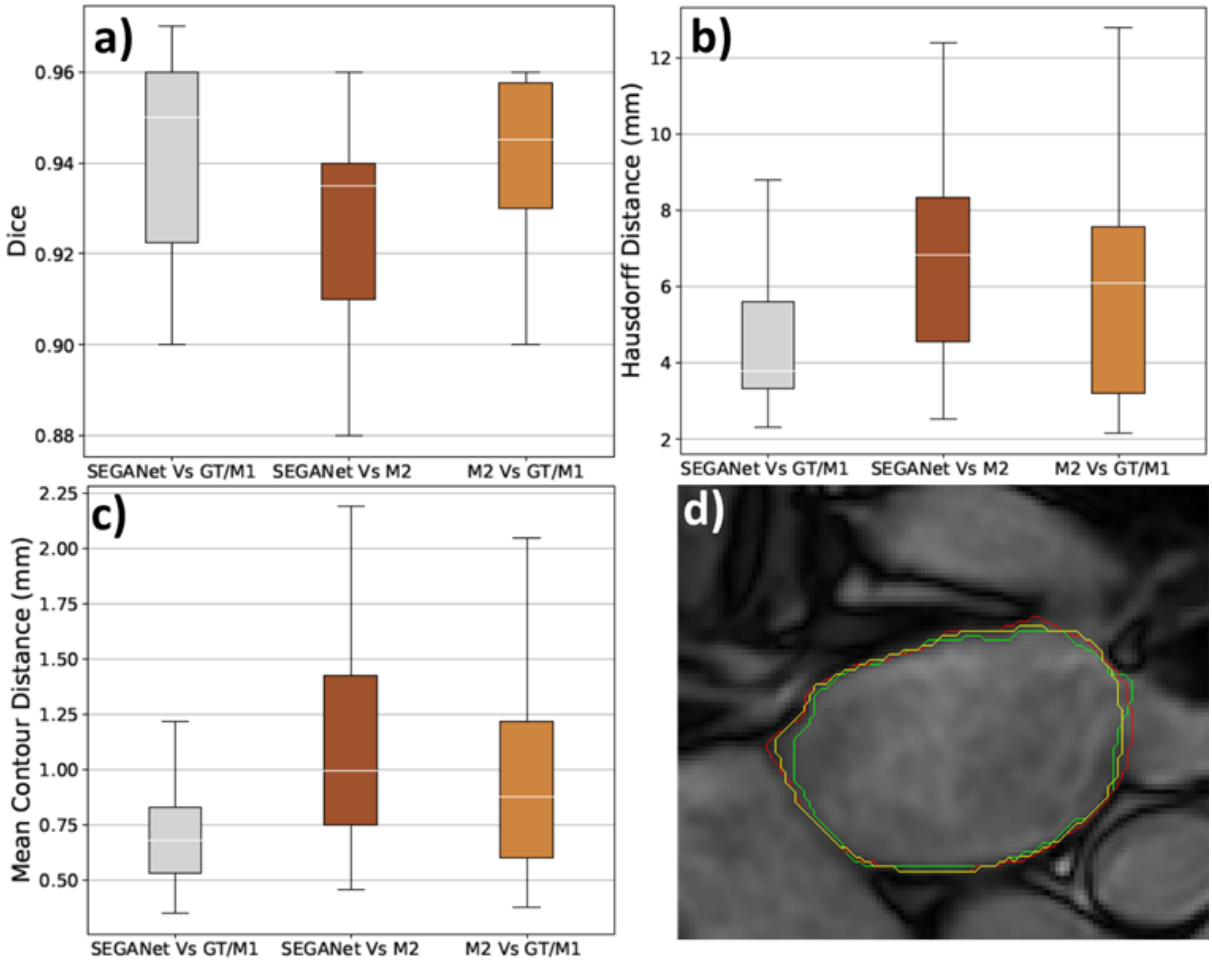

Figure 3.7: **a)** Dice coefficient (DC), **b)** Hausdorff distance (HD) and **c)** mean contour distance (MCD) comparing automatic (SEGANet) and manual (M2) LA segmentations with the ground truth (GT/M1) in 13 subjects. **d)** Contours indicating the results of the manual and automatic LA segmentations in a representative slice: yellow is the SEGANet automatic segmentation, red is the manual M2 segmentation, and green is the GT/M1. Image from [36].

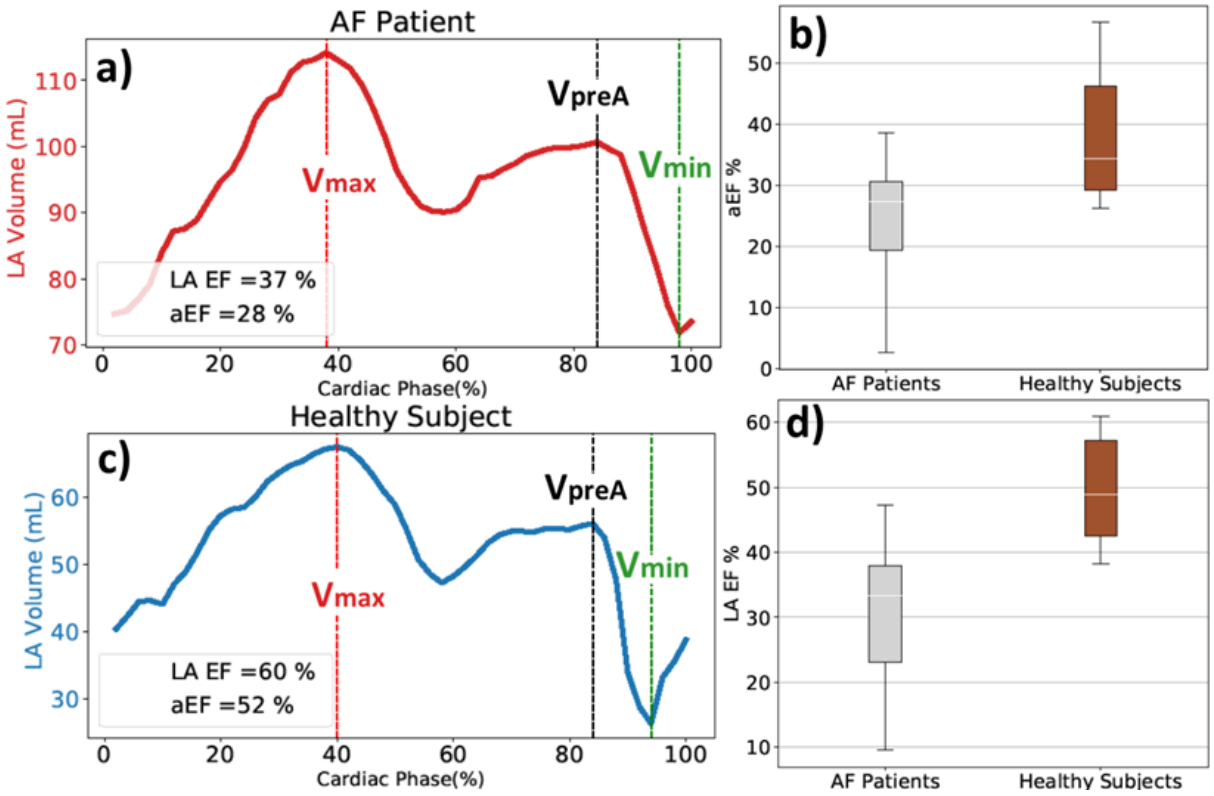

Figure 3.8: LA volume throughout the cardiac cycle for two representative subjects: **a)** an AF patient and **c)** an healthy subject. Three LA volumes are indicated: the maximal volume (V_{max}), the minimal volume (V_{min}) and the second peak volume, corresponding to the volume at the onset of the P wave, just before active atrial contraction (V_{preA}). LA EF and aEF values are shown in each subfigure. Moreover, box plots of **b)** aEF and **d)** LA EF for AF patients and healthy subjects show that values are significantly higher in healthy subjects than in patients (LA EF p-value: $< 1e^{-4}$; aEF p-value $< 1e^{-7}$).

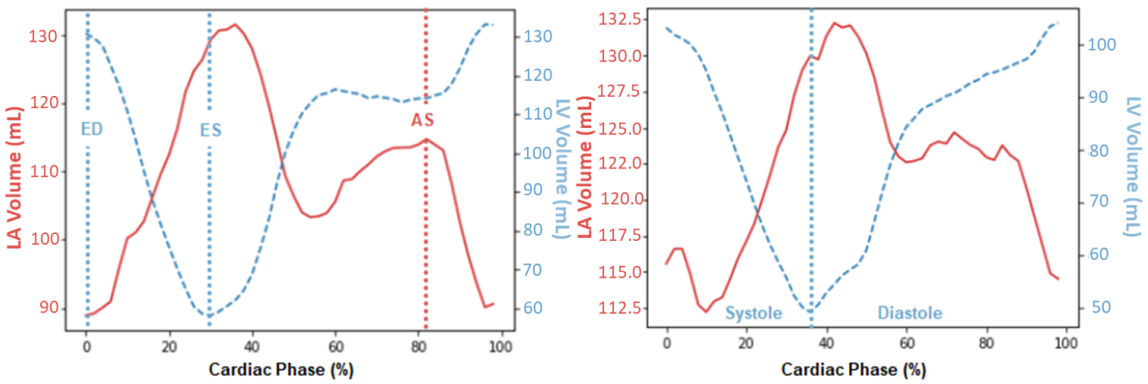

Figure 3.9: LA and LV volumes throughout the cardiac cycle for two representative subjects. Three phases of the cardiac cycle are indicated: end ventricular diastole (ED), end ventricular systole (ES) and atrial systole (AS). Image from [36].

3.3 DISCUSSION

SEGANet generated very good quality LA segmentations in short axis CINE MRI image stacks. The multi-slice CINE MRI segmentation was made on a slice-by-slice basis, which despite not making direct use of the 3D spatial and temporal correlations of the dataset, has other advantages [28, 108]. The first one is SEGANet's simplicity and low parameter numbers when compared to the 3D or 3D+time counterparts. The immunity to misregistrations across slices caused by variations in breath hold positions is also an advantage. Additionally, treating each slice independently allowed the SEGANet training process with comparatively fewer number of patient datasets than in 3D (or 3D+time) networks. Furthermore, since the training dataset had a relatively small size (715 slices), the extensive data augmentation strategy was crucial for the good SEGANet performance. Finally, this approach uniquely includes the LAA, that is important for assessing stroke risk in AF.

The MDC metric for SEGANet is on pair with the obtained values for 2D and 3D atrium segmentation from LGE MRI images (Tabel 3.3). However, CINE MRI images present a distinct contrast and resolution from LGE images, creating different challenges and thus, a direct analysis from segmentation results is consequently not ideal. Moreover, the estimated HD is also in accordance with the literature (Chen et al. 2018a [99] : 14.23 ± 4.83 mm). Looking at the segmentation quality metrics (Figure 3.7), the network works well in the regular and well-defined LA body, leading to small MCD values and high DC. Due to the morphological complexity and variability of the LAA, the contour distances (HD) tend to be relatively high.

Method	Description	MCD
Xiong et al. [34]	2D CNN	0.93 ± 0.005
Preetha et al. [85]	2D CNN	0.89
Bian et al. [86]	2D CNN	0.93
Chen et al. [99]	2D CNN	0.90 ± 0.93
Xia et al. [89]	Two-stage 3D segmentation framework	0.92 ± 0.007
Savioli et al. [88]	3D Volumetric-FCNN	0.77 ± 0.0787
Jia et al. [90]	3D U-Nets	0.92 ± 0.005
Vesal et al. [91]	3D CNN	0.93
Li et al. [92]	3D CNN	0.92
Mortazi et al. [93]	Multi-view 2D FCN	0.91
Yang et al. [94]	Multi-view 2D FCN	0.90 ± 0.053
SEGANet	2D CNN	0.93 ± 0.004

Table 3.3: MCD obtained with networks for 3D LGE images segmentation and SEGANet used in CINE MRI image stacks. Although all the NN presented the ability to segment LA, variations in contrast and resolution comparing CINE MRI and LGE images, created different challenges for SEGANet. Consequently, direct results analysis is not ideal.

Regarding ventricular automatic segmentation, the CNN performance is on par with other NNs for CINE-MRI [28, 42]. The method accurately identified all three labels when present and carried a short time penalty showing the network is able to generalize in multiple datasets.

Up until now, there were not DL-based approaches designed for LA segmentation from bSSFP CINE

MRI. The clinical potential of the technique is demonstrated when SEGANet was employed to obtain fully automatic estimates of LA volumes, LA EFs and aEFs. Currently, LA volumes and functional markers are estimated in CINE MRI by using empirical formulae to single-slice long axis views [39, 40], after manual annotation. The volumetric curves are in good agreement with the literature [109], showing the two characteristics peaks regarding the LA conduit and booster pump functions [19, 37].

The typical estimation of the atrial biomarkers from long axis 2D slices, unrealistically assume the LA shape is as geometrically regular as that of the LV (see Fig. 3.5), and rely on manual delineations of left atrial apex-base dimensions. Furthermore, due to the not isotropic atrial deformations across the cardiac cycle, it is not clear whether single 2D views allow for the accurate identification of the cardiac phases corresponding to maximal LA and minimal LA volume, as required for EF estimation. It is particularly difficult to identify the onset of atrial contraction (V_{preA}), required for aEF estimation, which our method performs automatically.

We expect that SEGANet will enhance the clinical usefulness of atrial CINE MRI imaging. The method should eventually provide information for AF treatment selection and make an important contribution to the stratification and treatment of AF patients. It is important to stress this is a highly pressing clinical problem due to the high incidence of AF and the relatively low success of available treatments.

4 AUTOMATIC MYOCARDIAL DISEASE PREDICTION

The EMIDEC STACOM 2020 challenge was used as a starting point to create reliable classification networks using as inputs a large dataset with static images and annotations per patient. For the challenge, NNs were designed and used to automatically predict myocardial disease from DE-CMR and clinical metadata, as detailed in `https://arxiv.org/abs/2010.08469`.

4.1 METHODOLOGY

4.1.1 DATASET: CLINICAL IMAGES AND METADATA

The inicial dataset comprised 100 anonymised cases: 67 pathological and 33 with normal findings, as detailed in [110, 111]. For each case, Phase Sensitive Inversion Recovery (PSIR) DE-CMR images covering the LV in 5-10 SA were made available, as well as the respective manual segmentations of 4 labels (Label 1: Blood Pool; Label 2: Normal myocardium; Label 3: MI; Label 4: No-reflow area). Moreover, 12 clinical variables were provided: sex, age, tobacco (Y/N/Former smoker), overweight (BMI: Body Mass Index > 25), arterial hypertension (Y/N), diabetes (Y/N), familial history of coronary artery disease (Y/N), ECG (ST+ (STEMI) or not), troponin (value), Killip max (1 - 4), EF of the LV from echography (value), and NTproBNP (value). Then, 50 additional cases, with the respective DE-CMR images and clinical information, were released during a live session to test the NNs.

4.1.2 IMAGE PREPROCESSING

Since the dataset had variable sizes in the z-axis, the DE-CMR images were zero-padded to obtain 10 slices in every input. Also, in order to eliminate anatomical structures unrelated to the heart, which could compromise the results, the images were cropped in-plane to a matrix of 128×128 using the LV blood pool segmentation from section 4.1.3. In detail, the blood pool label was employed to calculate a bounding box with the LV. Then, the centroid (x_c, y_c) was computed with a geometric decomposition method [112] described as:

$$x_c = \frac{\sum A_i x_{c_i}}{\sum A_i}, y_c = \frac{\sum A_i y_{c_i}}{\sum A_i} \tag{4.1}$$

where the shape is divided into a finite number of smaller areas (A_i) and the respective centroid (x_{ci}, y_{ci}) is estimated using standard formulae for simple configurations as detailed in [113].

The dataset was split into 3 smaller sets: training, validation and test (table 4.1). The distribution was done by randomly splitting the subjects from each class (myocardial disease and normal condition) into 6 smaller groups following a 70/10/20 ratio. Groups from distinct labels were then paired considering the same portion of cases (relative to the respective class). This division method avoids the presence of information from the same input in a different sets and allows a class balance across all three groups. Images were inputted into the network as:

$$T_I = (B \times S \times H \times W) \tag{4.2}$$

where B is the dataset length, S the number of slices, H the image height and W the width. In particular, for this task B varied along with the different sets (training, validation and test) and was used S

= 10, H = 128 and W = 128.

	Training Set	Validation Set	Test Set
Subjects	70	10	20
Slices	700	100	200
Percentages	70%	10%	20%

Table 4.1: Detailed data division of the 100 labelled cases.

Metadata was loaded as 2D tensors in the following configuration :

$$T_M = (S \times N) \tag{4.3}$$

where, S represents the dataset size and N indicates the number of metadata variables. For the challenge, S varied along with the different sets (training, validation and test) and N was always 12.

4.1.3 DE-CMR AUTOMATIC SEGMENTATION

The segmentation labels of the ventricles and scar structures were obtained with two NN-based methods. The approach was separately developed by other team members using the EMIDEC dataset. The first NN was trained with the Dice loss function to segment the LV blood pool region and used to crop each slice. The second NN was trained, using the generalized Dice loss function [114] and the cropped images as inputs, to automatically segment DE-CMR images into three categories (the LV myocardium, the LV blood pool and, where detected, the LGE uptake regions). In summary, the two NNs methods, both based on the 2D U-Net architecture [43, 104], composed an approach to automatically segment three labels in the DE-CMR images and then extract their volumes to use as extra inputs in the classification context.

4.1.4 DATA AUGMENTATION

Regarding the augmentation, multiple functions were applied, such as: 1) small image rotations and image flips; 2) additional cropping of the images; 3) additive stochastic noise; 4) k-space corruption; 5) smooth non-rigid deformations using free-form deformations and 6) intensity and contrast scaling. More details can be found in section 3.1.3.

4.1.5 NN FOR MYOCARDIAL DISEASE PREDICTION : CLINIC-NET AND DOC-NET

For the challenge, the following NNs were developed taking the presence or absence of myocardial disease as labels:

Clinic-NET: Classification NN based on clinical information. Clinic-NET is a classification NN with 3 fc layers based on clinical information only. The method encodes the 12-variable provided metadata in 3 layers with fc1 = 20 units, fc2 = 30 units, fc3 = 10 units and 2 final units (Figure 4.1). The network employed PReLu functions in the first three layer's outputs.

DOC-NET: Classification NN based on DE-CMR and other clinical information. The network for the DE-CMR images feature extraction contains seven convolutions layers as shown in Figure 4.2a, with 3D convolutions with $3\times3\times3$ kernels and a stride of 2. The channel dimensions of 4, 8, 16, 32, 64,

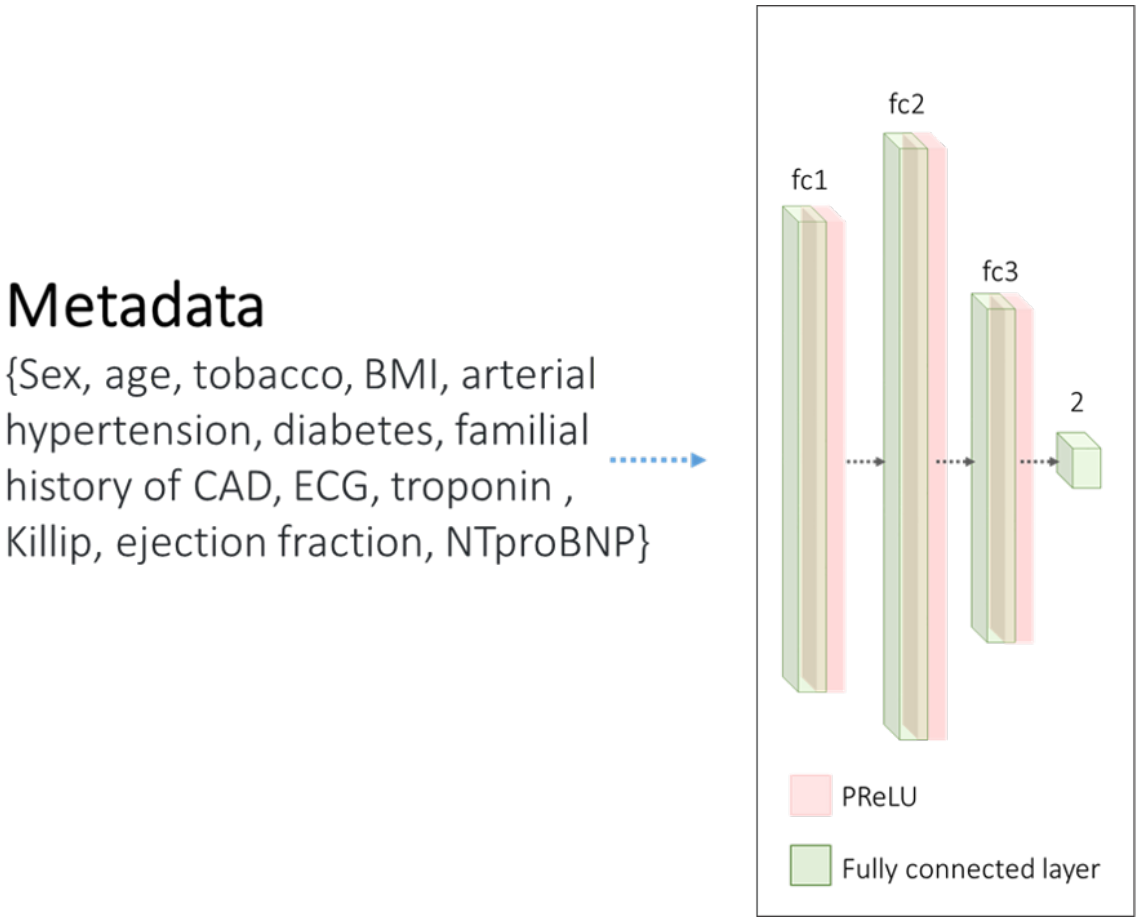

Figure 4.1: **Clinic-NET:** Classification NN based on clinical information composed with four fully connected (fc) layers.

16 and 8, firstly increase and then decrease the number of feature maps, improving the computational efficiency of the NN and ensuring both metadata and DE-CMR images (features) have a balanced contribution in the classification task. The image feature extraction network also includes instance layer-normalization, dropout (20% probability of being dropped out) and PReLU activation. After being flattened into an 8-element array, the image feature vector was concatenated with the 12 provided metadata variables, and taken as input into a fc-NN. The final network is similar to Clinic-NET (Figure 4.2b), however, to match the new input size, the 3 fc layers were rescaled to fc1 = 33 units, fc2 = 50 units, fc3 = 16 units.

Clinic-NET+ and DOC-NET+ use 16 metadata variables as inputs: volumetric information from previously segmented DE-CMR concatenated with the existing 12 metadata variables. Both NNs present the same structure as Clinic-NET and DOC-NET respectively.

The additional metadata used in Clinic-NET+ and DOC-NET+ were provided by the LV image segmentations from the ground truth data. The additional considered information was: LV blood pool volume; healthy myocardium volume; volume of LGE uptake area; volume of no-reflow area. Additional experiments were executed to gauge the importance of the volume information in the classification performance. In these experiments, Clinic-NET+ was trained and tested using these extra metadata variables only.

Since without manual segmentations it was not possible to deploy these NNs, segmentation networks were created by other team members to provide volumetric information in a stand-alone method, allowing classification in more general circumstances. The networks listed above (section 4.1.3), automatically

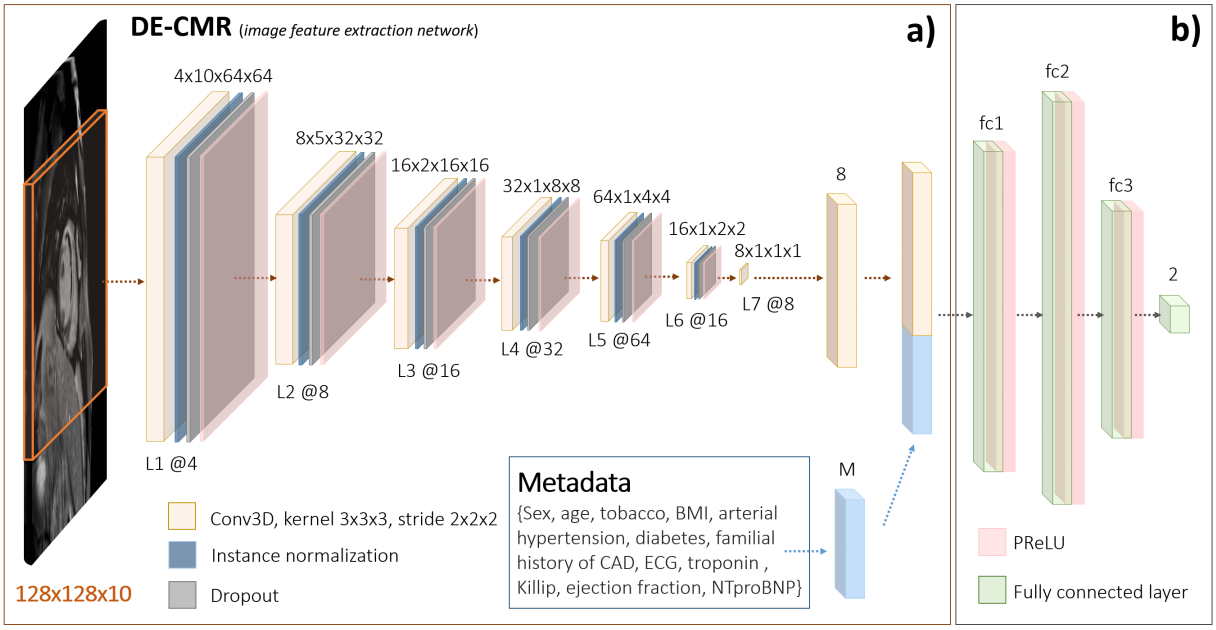

Figure 4.2: **DOC-NET classification network**: a) image feature vectors, obtained from the last convolutional layer of an image feature extraction network, are concatenated with the metadata vector (M) and b) sent through four fully connected (fc) layers. Image from [9].

segment DE-CMR into labels 1-3 and extract the volume of each class. Then, considering label 4 as 0 in all inputs, the pre-trained Clinic-NET+ and DOC-NET+ were tested with the new metadata variables. Results from testing these NNs with manual and automatic label volumes were compared using quality metrics as accuracy, sensitivity, specificity and confusion matrix.

Training Process For the classification, all four NNs (Clinic-NET, Clinic-NET+, DOC-NET, DOC-NET+) were trained throughout 170 iterations with all data in one batch. The CE loss function and the Adam optimizer with a learning rate of 0.00001 were employed. The method was implemented using Python (version 3.7) and PyTorch (version 1.7.0). The training process took 30-50 minutes using four K80 GPUs and sixteen CPUs.

Live Challenge The EMIDEC challenge required testing of the NNs during a live session using 50 additional test cases that were subsequently released at the time. Thus, Clinic-NET and Clinic-NET+ were retrained with the entire EMIDEC dataset maintaining the same details as above. For this purpose, the 100 labelled cases were redivided into training/validation with 80/20 ratio. Particularly, Clinic-NET+ was trained with label volumes provided by ground truth segmentations whereas during the live testing this was performed using volumetric information from automatic segmentation. In the end, the classification accuracies were calculated by the challenge organisers.

4.2 RESULTS

In the Emidec Stacom 2020 challenge, the proposed DE-CMR segmentation method provided qualitatively good results (Figure 4.3), allowing automatic crops in the region of interest and reliable estimation of LV, healthy myocardium and region of LGE uptake.

The overall classification performance reached a test accuracy of 100% in Clinic-NET+ and DOC-NET+, followed by DOC-NET (accuracy: 95%) and Clinic-NET (accuracy: 85%). Table 4.2 presents quality metrics (accuracy, sensitivity, specificity and confusion matrix (comprising the number of true

positive (TP), false positive (FP), false negative (FN) and true negative (TN) cases) obtained with all classification networks using the additional metadata extracted from the DE-CMR automatic segmentation. However, it is important to stress when tested with manual segmentation information the models achieved the same results.

In an additional experiment, Clinic-NET+ was trained with the extra label volumes only. The NN was tested with information from the stand-alone method (accuracy: 0.95 , sensitivity: 0.86 , specificity: 1, TN/FP/FN/TP: 13/0/1/6) and the results were similar when tested with volumetric labels extracted from manual segmentations (accuracy: 1; sensitivity: 1; specificity : 1; TN/FP/FN/TP: 13/0/ 0/7).

In the live challenge, the employed networks also performed well achieving high accuracy values (Clinic-NET : 72%, Clinic-NET+ : 82%) and rewarding the team with a second place.

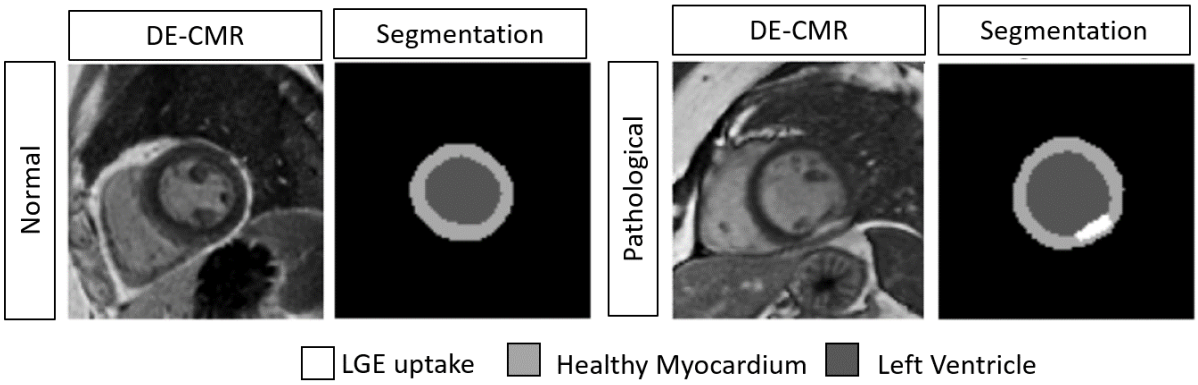

Figure 4.3: **DE-CMR images and segmentation** of the left ventricle (LV), normal myocardium, and region of LGE uptake (if present) obtained with the proposed automatic segmentation method for two slices from two representative subjects. Image from [9].

		Clinic-NET		DOC-NET		Clinic-NET+		DOC-NET+	
		Pathology	Normal	Pathology	Normal	Pathology	Normal	Pathology	Normal
Actual	Pathology	12	1	12	1	13	0	13	0
	Normal	2	5	0	7	0	7	0	7
Accuracy		0.85		0.95		1		1	
Sensitivity		0.92		0.92		1		1	
Specificity		0.71		1		1		1	

Table 4.2: **Confusion matrix** (actual vs predicted counts of pathological and normal cases), **accuracy, sensitivity, and specificity** obtained with all classification networks.

4.3 DISCUSSION

For the EMIDEC classification challenge a 4-layer fully-connected NN named Clinic-NET was initially proposed. The high classification performance of the network (accuracy: 85%) suggests that the clinical variables already provide essential classification information that can be leveraged by the method.

The prediction accuracy was improved when information from DE-CMR image segmentation, in both manual and automatic approaches, was added. The value of including volumetric information was confirmed in additional experiments where size of potential infarct areas was found the most important label followed by LV dimensions. In these trials, classification was also performed using only the

volumetric extra variables, achieving a performance comparable with Clinic-Net+ (100% accuracy). The reason for the outstanding performance of Clinic-NET+ and DOC-NET+ is probably related to a very high predictive value of the absence of the LGE zone segmentation in normal cases. Given the imbalance between the number of slices (10) and the in-plane matrix size (> 96 along each direction), it is possible that the performance of DOC-NET/DOC-NET+ may improve when using 2D rather than 3D convolutions to extract imaging features for the classification task.

For the live challenge, only two approaches could be presented, one with clinical information only and another with clinical information and DE-CMR images. Both adopted methods, Clinic-NET and Clinic-NET+ presented lower accuracy values (Clinic-NET : 72%, Clinic-NET+ : 82%) when compared with test metrics (Clinic-NET: 85%, Clinic-NET+: 100%), indicating that the trained NNs have low generalization capability. A metadata augmentation approach could help to reduce overfitting by increasing the dataset size. Adopting an unsupervised DL approach could also improve the results since it has started being investigated to optimize the model's performance [28]. Moreover, several images exhibited novel artefacts which potentially led to errors in automatic NN segmentation and the detection of false positives LGE uptake regions, influencing Clinic-NET+ results.

Regarding the automatic DE-CMR segmentation, this is particularly useful when ground truth segmentations are not available and the fact that the method does not segment the no-reflow area (label 4) the classification performance was not affected.

5 CLASSIFICATION OF AF RECURRENCE

For the final aim of the project, the classification NNs described in section 4 were re-trained to predict AF recurrence after CA in AF patients. For this task pre-ablation LA short axis in CINE MRI were used.

5.1 METHODOLOGY

5.1.1 DATASET: CLINICAL IMAGES AND METADATA

The dataset initially comprised short axis CINE MRI image stacks (acquired as detailed in section 3.1.1), however, only 40 samples also presented clinical variables including: sex, body mass index (BMI), heart rate, AF type (Paroxysmal/Persistent/Permanent), AF duration, LV EF, LA sphericity, LA maximum volume, LA EF, LA fibrosis score and other pathologies (Y/N). Thus, the 40 cases with the respective 11 metadata were considered for the task, 19 presenting disease recurrence within 1 year after ablation and 21 indicating stable recovery.

5.1.2 IMAGE PREPROCESSING

The 40 cases were split into training, validation and test sets (Table 5.1). From each scan, the slices covering the atria were automatically selected using the method described in section 3.1.5. For this task, 3 LA slices were then collected per patient. In detail, the chosen slices were the 3 most inferior (nearest to LV), which showed a larger area of LA and presented the LAA.

Since data acquisitions had variable number of cardiac phases, images were temporally interpolated to obtain 30 cycles. To remove anatomical structures unrelated to the heart, they were further cropped in-plane to a matrix of 128 x 128, whose centre was the LA segmentation centroid obtained with SEGANet (section 3). Metadata was computed as in 4.3 considering S as the minibatch size. Each slice was concatenated to the clinical information from the respective case and considered independent inputs.

The final image slices were loaded as follows:

$$T_M = (B \times T \times H \times W) \tag{5.1}$$

where, B represents the dataset length, T the number of cardiac phases, H the image height and W the width. For this task was used T = 30, H = 128 and W = 128.

	Training Set	Validation Set	Test Set
Subjects	28	4	8
Slices	84	24	12
Percentages	70%	10%	20%

Table 5.1: The distribution was done by randomly splitting the subjects from each class (recurrence and non-recurrence) into 6 smaller groups following the 70/10/20 ratio. Groups from distinct labels were then paired considering the same portion of cases (relative to the respective class). This division method avoids the presence of information from the same input in different sets and allows a class balance across all three groups. After division, 3 LA slices from each case were processed and repeatedly concatenated to clinical information regarding the sample.

5.1.3 DATA AUGMENTATION

Image augmentation was applied as in the Emidec challenge (secction 4) and metadata augmentation was introduced in training and validation sets. For this, a variable was replaced each time by the mean of the same variables for the remaining subjects. This is inspired by the mean inputation method used to deal with missing data. This augmentation procedure leads to metadata inputs V times bigger, where V is the number of variables considered.

5.1.4 NN FOR AF RECURRENCE: CLINIC-NET AND DOC-NET

The classification task used the following NNs to predict AF recurrence/non-recurrence after CA in AF patients:

Clinic-NET: Classification NN based on clinical information. Clinic-NET was re-trained with the 11 provided metadata variables, firstly using 28 clinical inputs and then with 336 augmented information. The structure was mantained as above (Figure 4.2b), except for some training process parameters, such as the learning rate and regularization, which were optimized. The most accurate results were obtained with augmented data, a learning rate of 0.001 , 20400 iterations and a weight decay of 0.01 (L2 regularization).

DOC-NET: Classification NN based on DE-CMR and other clinical information. The network accepted as input a single slice covering 30 cardiac phases and connected with the respective clinical data. In detail, DOC-NET combined features extracted from CINE-MRI stacks and features computed from metadata to perform the prediction. During the training process optimization was executed kepping DOC-NET structure (Figure 4.2a) as well as the augmentation referred above in the 84 input tensors. The most accurate results were obtained with 1700 iterations, a learning rate of 0.001 and a dropout of 80% (80% probability of being dropped out).

Clinic-NET+ To further explore the classification task, extra metadata variables were created with volumes of the LA automatic segmentation label (section 3) and concatenated with the existing metadata. In detail, the LA volumes throughout the cardiac cycle (one per phase) were computed from SEGANet segmentations, producing arrays with the same dimension as the images cardiac phases (between 30 and 50 as detailed in section 3.1.1). The outputs were cropped to generate arrays all with the same size, producing 25-element arrays. Thus, Clinic-NET+ was re-trained with 36 metadata variables (the initial 11 + the extra 25 inputs), using both 46 inputs (the initial 11-element arrays augmented) and 1000 augmented data (the 36-element arrays augmented). Maintaining Clinic-NET+ structure, the best accuracies were obtained with augmented data, a learning rate of 0.00001, 20400 iterations and and mini-batches of 500.

Training Process All the NN were trained using Python (version 3.7) and PyTorch (version 1.7.0). The training process ranged from 10-30 minutes, using four K80 GPUs and sixteen CPUs. The models were evaluated using quality metrics as accuracy, sensitivity, specificity and confusion matrix.

5.2 RESULTS

For the prediction of AF recurrence, after parameter fine tuning, the best overall performance was jointly achieved by Clinic-NET+ with a test accuracy of 88% followed by Clinic-NET (accuracy: 75%) and DOC-NET (accuracy: 63%) - see Table 5.2. However, the results showed some variability when networks where retrained and tested on different subsets.

		Clinic-NET		DOC-NET		Clinic-NET+	
		AF Recurrence	Stable Recovery	AF Recurrence	Stable Recovery	AF Recurrence	Stable Recovery
Actual	AF Recurrence	4	0	9	3	4	0
	Stable Recovery	2	2	6	6	1	3
	Accuracy	0.75		0.63		0.88	
	Sensitivity	1		0.75		0.75	
	Specificity	0.5		0.5		1	

Table 5.2: **Confusion matrix** (actual vs predicted counts of AF recurrence and non-recurrence cases), **accuracy, sensitivity, and specificity** obtained with all re-trained classification networks. DOC-NET accepted 3 LA slices per patient, using 12 inputs in the test set. Since CLINIC-NET and CLINIC-NET+ used one input tensor per subject, both test sets only comprised 8 cases.

5.3 DISCUSSION

In the prediction of AF recurrence after CA, the networks' performance achieved less accurate results when compared with the EMIDEC Stacom 2020 (Table 4.2). The parameters that more influenced the results had shown to be, during optimization, the number of iterations and the presence of augmentation. The small amount of available data is probably one of the reasons for the lack of generalization capacity.

More data could reduce overfitting, introduce more physiological variability and improve the prediction performance. Also, the AF recurrence prediction is a more difficult problem where the features of interest are more subtly presented in the data. In particular, DOC-NET that used image inputs had shown the lowest results. In detail, in this method from each scan, 3 LA slices were collated and concatenated to the same clinical information. Thus, the results suggested the use of repeated equal metadata increases the overfitting.

In order to optimize the networks performance, future plans could include an unsupervised DL approach [28] and other MRI techniques as DE-CMR images.

6 CONCLUSION

This project aimed to create a suite of DL-based tools to automatically post-process cardiac MRI and extract relevant clinical information from medical images.

For the segmentation task SEGANet was presented for fully automatic segmentation of the LA from CINE MRI (section 3). The provided SA CINE MRI image stacks with full cardiac coverage were the first key factor for project success. The network led to accurate segmentations of LA across the cardiac cycle, on par with other NNs [34, 94]. Comparing with 3D (or 3D+time) networks, SEGANet presented high performance using a smaller dataset and low parameter numbers. Additionally, this method is not affected to misregistrations across slices caused by variations in breath-hold positions, and also uniquely incorporates the LAA in segmentations, which is relevant for assessing stroke risk in AF. Since DL-based methods tend to be biased by their training datasets, generalization capabilities are usually a limitation of these models. Thus, testing SEGANet in different datasets, such as with data from different scanners and in patients with other pathologies, could be a crucial next step to evaluate the clinical usability of the model.

In the classification of myocardial disease from DE-CMR and patient clinical information (section 4) the NNs presented high accuracy results. The 3D methods achieved a performance on par with the methods exhibited on the live challenge, rewarding the team with a second place. Employ metadata augmentation would be useful future steps to further enhance the methods' performance. Also, validating the models in a larger number of cases, including patients with undetected MI on DE-CMR would help measure clinical viability.

Finally, the prediction of AF recurrence after CAs (section 5) achieved less accurate results, when compared with the myocardial disease prediction. However, direct comparisons between the classification networks are challenging and not reliable since the performance depends on the dataset and the task. Classifications NNs are not developed as segmentation networks, few clinical biomarkers have been studied/predicted. This method introduces a novel and automatic technique to analyse AF recurrence taking direct advantage from cardiac MRI images. Moreover, it offers higher simplicity when compared to manual techniques (described in section 2.5). Future steps would include methods to reduce overfitting and create more accurate results.

In this project, important steps have been taken to produce clinical benefits with DL applied in cardiac MRI. All this work showed the potential of NNs to interpret and extract clinical information from medical images. In the future, as more data becomes available, these methods will be able to help and guide clinical AF prognosis and diagnosis.

REFERENCES

[1] F. M. Filipoiu, *Atlas of Heart Anatomy and Development*, online ed. Springer, 2014.

[2] P. A. Iaizzo, *Handbook of Cardiac Anatomy, Physiology, and Devices*, 3rd ed. Totowa: Humana Press, 2006.

[3] R. Beigel, N. C. Wunderlich, S. Y. Ho, R. Arsanjani, and R. J. Siegel, "The left atrial appendage: Anatomy, function, and noninvasive evaluation," *JACC: Cardiovascular Imaging*, vol. 7, no. 12, pp. 1251–1265, 2014.

[4] H. Thomas et al, "Global atlas of cardiovascular disease 2000-2016: The path to prevention and control," *Global Heart*, vol. 13, no. 3, pp. 143–163, 2018.

[5] World Health Organization, "Cardiovascular diseases," Available at https://www.who.int/health-topics/cardiovascular-diseases (2020/12/06).

[6] A. N. Ganesan et al, "Long-term outcomes of catheter ablation of atrial fibrillation: a systematic review and meta-analysis," *Journal of the American Heart Association*, vol. 2, no. 2, p. e004549, 2013.

[7] M. Varela et al, "Novel computational analysis of left atrial anatomy improves prediction of atrial fibrillation recurrence after ablation," *Frontiers in Physiology*, vol. 8, p. 68, 2017.

[8] J. T. Kowallick et al, "Quantification of atrial dynamics using cardiovascular magnetic resonance: inter-study reproducibility," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, no. 1, p. 36, 2015.

[9] A. Lourenço et al, "Automatic myocardial disease prediction from delayed-enhancement cardiac mri and clinical information," *arXiv: 2010.08469*, 2020.

[10] A. E. Arai, "The cardiac magnetic resonance (CMR) approach to assessing myocardial viability," *Journal of Nuclear Cardiology*, vol. 18, no. 6, pp. 1095–1102, 2011.

[11] N. Bettencourt, A. Chiribiri, A. Schuster, and E. Nagel, "Assessment of myocardial ischemia and viability using cardiac magnetic resonance," *Current heart failure reports*, vol. 6, no. 3, pp. 142–153, 2009.

[12] C. M. Kramer and R. C. Heede, "Role of Cardiac MR Imaging in Cardiomyopathies," *Journal of nuclear medicine*, vol. 56, no. Suppl 4, pp. 39S–45S, 2015.

[13] R. O. Bonow et al, "Myocardial viability and survival in ischemic left ventricular dysfunction," *New England Journal of Medicine*, vol. 364, no. 17, pp. 1617–1625, 2011.

[14] R. Bailey, "The heart wall is made up of 3 layers that have their own functions," Available at https://www.thoughtco.com/the-heart-wall-4022792 (2021/01/16).

[15] A. L. Mescher and L. C. Junqueira, *Junqueira's Basic Histology: Text and Atlas*, 15th ed. New York: McGraw-Hill Education, 2016.

[16] M. T. Ashworth, "Chapter 19 – the cardiovascular system," in *Keeling's Fetal and Neonatal Pathology,* 5th ed. Cham: Springer, 2015, pp. 481– 529.

[17] T. E. of Encyclopaedia Britannica, "Systole," Available at https://www.britannica.com/science/systole-heart-function (2020/11/10).

[18] S. Nattel, "New ideas about atrial fibrillation 50 years on," *Nature*, vol. 415, no. 6868, pp. 219–226, 2002.

[19] B. D. Hoit, "Left atrial size and function: role in prognosis," *Journal of the American College of Cardiology*, vol. 63, no. 6, pp. 493–505, 2014.

[20] M. K. Konings et al, "A new electric method for non-invasive continuous monitoring of stroke volume and ventricular volume-time curves," *Biomedical engineering online*, vol. 11, no. 1, pp. 1–29, 2012.

[21] Y. Hagiwara et al, "Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review," *Information Sciences*, vol. 467, pp. 481–529, 2018.

[22] S. Leighton, "Catheter ablation for afib," Available at https://www.health.harvard.edu/heart-health/zap-away-atrial-fibrillation (2021/01/01).

[23] P. J. Lynch, "Drawing showing anterior left ventricle wall infarction," Available at https://en.wikipedia.org/wiki/Myocardial_infarction#/media/File:Heart_ant_wall_infarction.jpg (2020/10/10).

[24] D. Broadbent, A. Kidambi, and J. Biglands, "Cardiovascular magnetic resonance physics for clinicians," Available at https://www.escardio.org/static-file/Escardio/Subspecialty/EACVI/CMR%20Physics%20Pocket%20Guide%20iBook%20v1.0.pdf (2020/07/10).

[25] S. Uribe et al, "Whole-heart cine mri using real-time respiratory self-gating," *Magnetic Resonance in Medicine*, vol. 57, no. 3, pp. 606–613, 2007.

[26] C. B. Marcu, A. M. Beek, and A. C. Van Rossum, "Clinical applications of cardiovascular magnetic resonance imaging," *Canadian Medical Association Journal*, vol. 175, no. 8, pp. 911–917, 2006.

[27] S. Lynch, *Dynamical Systems with Applications Using Mathematica*, 2nd ed. Cham: Springer, 2017.

[28] C. Chen et al, "Deep learning for cardiac image segmentation: A review," *Frontiers in Cardiovascular Medicine*, vol. 7, p. 25, 2020.

[29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, online ed. MIT Press, 2016.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015, pp. 234–241.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2016, pp. 770–778.

[32] D. Nautiyal, "Underfitting and Overfitting in Machine Learning," Available at https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/ (2020/11/07).

[33] Baeldung, "How ReLU and Dropout Layers Work in CNNs," Available at https://www.baeldung.com/cs/ml-relu-dropout-layers (2020/11/07).

[34] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 515–524, 2019.

[35] G. A. Bello et al, "Deep-learning cardiac motion analysis for human survival prediction," *Nature Machine Intelligence*, vol. 1, no. 2, pp. 95–104, 2019.

[36] A. Lourenço et al, "Left atrial ejection fraction estimation using seganet for fully automated segmentation of cine mri," *arXiv: 2008.13718*, 2020.

[37] B. D. Hoit, "Evaluation of left atrial function: current status," *Structural Heart*, vol. 1, no. 3-4, pp. 109–120, 2017.

[38] A. K. Attili, A. Schuster, E. Nagel, J. H. Reiber, and R. J. van der Geest, "Quantification in cardiac mri: advances in image acquisition and processing," *The international journal of cardiovascular imaging*, vol. 26, no. 1, pp. 27–40, 2010.

[39] R. Erbel et al, "Comparison of single-plane and biplane volume determination by two-dimensional echocardiography 1. asymmetric model hearts," *European Heart Journal*, vol. 3, no. 5, pp. 469–480, 1982.

[40] K. Ujino et al, "Two-dimensional echocardiographic methods for assessment of left atrial volume," *The American journal of cardiology*, vol. 98, no. 9, pp. 1185–1188, 2006.

[41] J. Dernellis and M. Panaretou, "Left atrial function in patients with a high c-reactive protein level and paroxysmal atrial fibrillation," *Acta cardiologica*, vol. 61, no. 5, pp. 507–511, 2006.

[42] W. Bai et al, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, p. 65, 2018.

[43] A. Lourenço et al, "Left atrial ejection fraction estimation using SEGANet for fully automated segmentation of CINE MRI," *arXiv: 2008.13718*, 2020.

[44] C. J. Soriano, F. Ridocci, J. Estornell, J. Jimenez, V. Martinez, and J. A. De Velasco, "Noninvasive diagnosis of coronary artery disease in patients with heart failure and systolic dysfunction of uncertain etiology, using late gadolinium-enhanced cardiovascular magnetic resonance," *Journal of the American College of Cardiology*, vol. 45, no. 5, pp. 743–748, 2005.

[45] E. Lee, E.-S. H. Ibrahim, P. Parwani, N. Bhave, and J. Stojanovska, "Practical guide to evaluating myocardial disease by cardiac MRI," *American Journal of Roentgenology*, vol. 214, no. 3, pp. 546–556, 2020.

[46] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography," *Journal of the American College of Cardiology*, vol. 68, no. 21, pp. 2287–2295, 2016.

[47] B. Baeßler et al, "Mapping tissue inhomogeneity in acute myocarditis: a novel analytical approach to quantitative myocardial edema imaging by T2-mapping," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, no. 1, pp. 1–11, 2015.

[48] J. Mantilla et al, "Detection of fibrosis in late gadolinium enhancement cardiac mri using kernel dictionary learning-based clustering," in *2015 Computing in Cardiology Conference*. IEEE, 2015, pp. 357–360.

[49] T. Leiner et al, "Machine learning in cardiovascular magnetic resonance: basic concepts and applications," *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1, pp. 1–14, 2019.

[50] R. Rubin, D. S. Strayer, and E. Rubin, *Rubin's pathology: clinicopathologic foundations of medicine*, 6th ed. Philadelphia: Lippincott Williams & Wilkins, 2012.

[51] E. K. K. Brakohiapa, B. O. Botwe, B. D. Sarkodie, E. K. Ofori, and J. Coleman, "Radiographic determination of cardiomegaly using cardiothoracic ratio and transverse cardiac diameter: can one size fit all? part one." *The Pan African medical journal*, vol. 27, no. 201, 2017.

[52] A. Bouzas-Mosquera et al, "Left atrial size and risk for all-cause mortality and ischemic stroke," *Canadian Medical Association Journal*, vol. 183, no. 10, pp. E657–E664, 2011.

[53] N. Kawel et al, "Middle-aged and older subjects with steady-state free precession cardiac magnetic resonance the multi-ethnic study of atherosclerosis," *Circulation: Cardiovascular Imaging*, vol. 5, no. 4, pp. 500–508, 2012.

[54] S. Standring, *Gray's anatomy: the anatomical basis of clinical practice*, 41st ed., ser. Gray's Anatomy. New York: Elsevier, 2016.

[55] Z. A. A. İkiz, H. Üçerler, and T. Özgür, "Anatomic characteristics of left atrium and openings of pulmonary veins," *Anadolu Kardiyol Derg*, vol. 14, no. 8, pp. 674–678, 2014.

[56] S. Meek and F. Morris, "Introduction. I—Leads, rate, rhythm, and cardiac axis," *British Medical Journal*, vol. 324, no. 7334, pp. 415–418, 2002.

[57] P. Kirchhof et al, "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal*, vol. 37, no. 38, pp. 2893–2962, 2016.

[58] S. Thanigaimani et al, "Progression and reversibility of stretch induced atrial remodeling: Characterization and clinical implications," *Progress in Biophysics and Molecular Biology*, vol. 130, pp. 376–386, 2017.

[59] H. Calkins et al, "2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design: a report of the Heart Rhythm Society (HRS) Task Force on Catheter and Surgical Ablation of Atrial Fibrillation. Developed in partnership with the European Heart Rhythm Association (EHRA), a registered branch of the European Society of Cardiology (ESC) and ..." *Europace*, vol. 14, no. 4, pp. 528–606, 2012.

[60] A. Berruezo et al, "Pre-procedural predictors of atrial fibrillation recurrence after circumferential pulmonary vein ablation," *European Heart Journal*, vol. 28, no. 7, pp. 836–841, 2007.

[61] F. Bisbal et al, "Left atrial sphericity: a new method to assess atrial remodeling. impact on the outcome of atrial fibrillation ablation," *Journal of cardiovascular electrophysiology*, vol. 24, no. 7, pp. 752–759, 2013.

[62] G. E. Supple, "Can left atrial function guide atrial fibrillation ablation?" *Heart Rhythm*, vol. 12, no. 1, pp. 19–20, 2015.

[63] G. G. Blume et al, "Left atrial function: physiology, assessment, and clinical implications," *European Journal of Echocardiography*, vol. 12, no. 6, pp. 421–430, 2011.

[64] J. C. Jayaraj, K. Davatyan, S. Subramanian, and J. Priya, "Chapter 2 – epidemiology of myocardial infarction," in *Myocardial Infarction,* online ed. IntechOpen, 2019, pp. 10 – 19.

[65] S. H. Ralston and R. Britton, *Davidson's principles and practice of medicine*, 23rd ed. Edinburgh: Elsevier, 2018.

[66] A. K. Luo and K. C. Wu, "Imaging microvascular obstruction and its clinical significance following acute myocardial infarction," *Heart Failure Reviews*, vol. 11, no. 4, pp. 305–312, 2006.

[67] A. J. Perricone and R. S. Vander Heide, "Novel therapeutic strategies for ischemic heart disease," *Pharmacological Research*, vol. 89, no. 2, pp. 36–45, 2014.

[68] S. G. Myerson, J. Francis, and S. Neubauer, *Cardiovascular Magnetic Resonance*, ser. Oxford Specialist Handbooks in Cardiology. New York: Oxford University Press, 2010.

[69] M. Salerno et al, "Recent advances in cardiovascular magnetic resonance: Techniques and applications," *Circulation: Cardiovascular imaging*, vol. 10, no. 6, p. e003951, 2017.

[70] M. S. Nacif et al, "Cardiac magnetic resonance imaging and its electrocardiographs (ECG): tips and tricks," *International Journal of Cardiovascular Imaging*, vol. 28, no. 6, pp. 1465–1475, 2012.

[71] R. A. C. V. Nagel, E. and E. Fleck, *Cardiovascular magnetic resonance*, online ed. Steinkopff, 2004.

[72] K. Scheffler and S. Lehnhardt, "Principles and applications of balanced SSFP techniques," *European Radiology*, vol. 13, no. 11, pp. 2409–2418, 2003.

[73] N. Bettencourt, A. Chiribiri, A. Schuster, and E. Nagel, "Assessment of myocardial ischemia and viability using cardiac magnetic resonance," *Current Heart Failure Reports*, vol. 6, no. 3, pp. 142–153, 2009.

[74] J. W. Weinsaft, I. Klem, and R. M. Judd, "Mri for the assessment of myocardial viability," *Magnetic resonance imaging clinics of North America*, vol. 15, no. 4, pp. 505–525, 2007.

[75] R. J. Kim et al, "The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction," *New England Journal of Medicine*, vol. 343, no. 20, pp. 1445–1453, 2000.

[76] B. L. Gerber et al, "Prognostic value of myocardial viability by delayed-enhanced magnetic resonance in patients with coronary artery disease and low ejection fraction: impact of revascularization therapy," *Journal of the American College of Cardiology*, vol. 59, no. 9, pp. 825–835, 2012.

[77] K. C. Allman, L. J. Shaw, R. Hachamovitch, and J. E. Udelson, "Myocardial viability testing and impact of revascularization on prognosis in patients with coronary artery disease and left ventricular dysfunction: a meta-analysis," *Journal of the American College of Cardiology*, vol. 39, no. 7, pp. 1151–1158, 2002.

[78] M. Petrou and C. Petrou, "Chapter 6 – Image segmentation and edge detection," in *Image Processing: The Fundamentals,* 2nd ed. Hoboken: Wiley, 2011, pp. 542–544.

[79] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. e158–e177, 2011.

[80] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv: 1505.00853*, 2015.

[81] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *arXiv: 1502.01852*, 2015.

[82] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2017.

[83] X.-Y. Zhou, P. Li, Z.-Y. Wang, and G.-Z. Yang, "U-net training with instance-layer normalization," in *International Workshop on Multiscale Multimodal Medical Imaging*. Cham: Springer, 2019, pp. 101–108.

[84] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, 1978.

[85] C. J. Preetha, S. Haridasan, V. Abdi, and S. Engelhardt, "Segmentation of the left atrium from 3D gadolinium-enhanced MR images with convolutional neural networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer, 2018, pp. 265–272.

[86] C. Bian et al, "Pyramid network with online hard example mining for accurate left atrium segmentation," in *international workshop on statistical atlases and computational models of the heart*. Cham: Springer, 2018, pp. 237–245.

[87] C. Li, M. Z. Zia, Q.-H. Tran, X. Yu, G. D. Hager, and M. Chandraker, "Deep supervision with intermediate concepts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1828–1843, 2018.

[88] N. Savioli, G. Montana, and P. Lamata, "V-FCNN: volumetric fully convolution neural network for automatic atrial segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer, 2018, pp. 273–281.

[89] Q. Xia, Y. Yao, Z. Hu, and A. Hao, "Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer, 2018, pp. 211–220.

[90] S. Jia et al, "Automatically segmenting the left atrium from cardiac images using successive 3D U-nets and a contour loss," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer, 2018, pp. 221–229.

[91] S. Vesal, N. Ravikumar, and A. Maier, "Dilated convolutions in neural networks for left atrial segmentation in 3d gadolinium enhanced-mri," in *International workshop on statistical atlases and computational models of the heart*. Cham: Springer, 2018, pp. 319–328.

[92] C. Li et al, "Attention based hierarchical aggregation network for 3D left atrial segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer, 2018, pp. 255–264.

[93] A. Mortazi, R. Karim, K. Rhode, J. Burt, and U. Bagci, "CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2017, pp. 377–385.

[94] 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 *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2018, pp. 1123–1127.

[95] A. Esteva et al, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[96] V. Gulshan et al, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Journal of the American Medical Association*, vol. 316, no. 22, pp. 2402–2410, 2016.

[97] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2016, pp. 2818–2826.

[98] D. Peressutti et al, "A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction," *Medical image analysis*, vol. 35, pp. 669–684, 2017.

[99] C. Chen, W. Bai, and D. Rueckert, "Multi-task Learning for Left Atrial Segmentation on GE-MRI," in *International workshop on statistical atlases and computational models of the heart*. Cham: Springer, 2018, pp. 292–301.

[100] S. E. Petersen et al, "UK Biobank's cardiovascular magnetic resonance protocol," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, pp. 1–7, 2015.

[101] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Cham: Springer, 2016, pp. 630 – 645.

[102] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297-302, 1945.

[103] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv: 1412.6980*, 2014.

[104] E. Kerfoot et al, "Automated CNN-Based Reconstruction of Short-Axis Cardiac MR Sequence from Real-Time Image Data," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Cham: Springer, 2018, pp. 32 – 41.

[105] M. Mika et al, "Impact of reduced left atrial functions on diagnosis of paroxysmal atrial fibrillation: Results from analysis of time-left atrial volume curve determined by two-dimensional speckle tracking," *Journal of Cardiology*, vol. 57, no. 1, pp. 89–94, 2011.

[106] H. chubb et al, "The reproducibility of late gadolinium enhancement cardiovascular magnetic resonance imaging of post-ablation atrial scar: a cross-over study," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, pp. 1–15, 2018.

[107] D. Victoria et al, "Fate of left atrial function as determined by real-time three-dimensional echocardiography study after radiofrequency catheter ablation for the treatment of atrial fibrillation," *The American Journal of Cardiology*, vol. 101, no. 9, pp. 1285 – 1290, 2008.

[108] C. Chen, C. Biffi, G. Tarroni, S. E. Petersen, W. Bai, and D. Rueckert, " Learning Shape Priors for Robust Cardiac MR Segmentation from Multi-view Images," *arXiv: 1907.09983*, 2019.

[109] V. Järvinen, M. Kupari, P. Hekali, and V.-P. Poutanen, "Assessment of left atrial volumes and phasic function using cine magnetic resonance imaging in normal subjects," *The American journal of cardiology*, vol. 73, no. 15, pp. 1135–1138, 1994.

[110] EMIDEC, "EMIDEC classification challenge:," Available at http:/emidec.com/classification-contest (2020/07/06).

[111] A. Lalande et al, "Emidec: A Database Usable for the Automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI," *Data*, vol. 5, no. 4, p. 89, 2020.

[112] S. Ibrahim, "Lecture notes: Centroid and centre of mass of composite bodies," Available at https://www.researchgate.net/publication/307605237_Centroid_and_Centre_of_mass_of_composite_bodies (2021/05/01).

[113] M. E. Lemonis, "Centroids - reference table," Available at https://calcresource.com/centroids-list.html (2021/05/01).

[114] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "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*. Cham: Springer, 2017, pp. 240–248.