Deep learning on large clinical databases for image-based predictions of cardiac arrhythmias

Buntheng Ly

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THÈSE DE DOCTORAT

Apprentissage profond sur de grandes bases de données cliniques pour la prédiction des arythmies cardiaques à partir de l'imagerie

Buntheng LY

INRIA, Équipe EPIONE

LIRYC, Équipe MULTIMODAL DATA SCIENCE

Thèse dirigée par Maxime SERMESANT et Hubert COCHET

Soutenue le 20 Décembre 2022

Présentée en vue de l'obtention du grade de DOCTEUR EN AUTOMATIQUE, TRAITEMENT DU SIGNAL ET DES IMAGES de l'UNIVERSITÉ CÔTE D'AZUR.

Devant le jury composé de :

Eloi MARIJON	AP-HP, 3IA PRAIRIE	Rapporteur
Ender KONUKOGLU	ETH-Zurich	Rapporteur
Pierre JAÏS	Université de Bordeaux, Liryc	Président
Hubert COCHET	Université de Bordeaux, Liryc	Directeur de thèse
Maxime SERMESANT	Inria Sophia Antipolis, Liryc	Directeur de thèse

Epione e-patient / e-medicine

Inria

liryc ELECTROPHYSIOLOGY AND HEART MODELING INSTITUTE

Membres du Jury

Titre français
Apprentissage profond sur de grandes bases de données cliniques pour la prédiction des arythmies cardiaques à partir de l'imagerie

Titre anglais
Deep Learning on Large Clinical Databases for Image-based Predictions of Cardiac Arrhythmias

Devant le jury composé de :

Président du jury
Pierre JAÏS, Professeur, Université de Bordeaux, Liryc

Rapporteurs

Eloi MARIJON, Professeur, AP-HP, 3IA PRAIRIE
Ender KONUKOGLU, Professeur associé, ETH-Zurich

Examinateurs

Eloi MARIJON, Professeur, AP-HP, 3IA PRAIRIE
Ender KONUKOGLU, Professeur associé, ETH-Zurich
Hubert COCHET, Professeur, Université de Bordeaux, Liryc
Maxime SERMESANT, Directeur de recherche - HDR, Inria Sophia Antipolis, Liryc

Résumé

La mort subite (MS) est un problème de santé majeur, responsable d'une mortalité annuelle de 400 000 personnes aux États-Unis et d'un chiffre similaire en Europe. La mort subite est la conséquence de l'arrêt soudain et total de l'activité cardiaque, également connu sous le nom d'arrêt cardiaque soudain (ACS). Une étude de surveillance de l'activité cardiaque a établi un lien entre l'ACS et l'arythmie ventriculaire (AV), qui est responsable de 70% de tous les ACS enregistrés. L'approche préventive de l'arythmie ventriculaire est le défibrillateur automatique implantable (DAI), tandis que l'ablation par radiofréquence (AR) est une thérapie potentiellement curative. Le DAI surveille l'activité électrique cardiaque et est programmé pour délivrer des chocs électriques afin d'arrêter l'épisode d'AV. Cependant, les critères de sélection des patients, souvent basés sur l'indice de la fraction d'éjection du ventricule gauche (FEVG), constituent une limite importante, avec un grand nombre de faux positifs et négatifs. La AR est une procédure interventionnelle qui permet d'éliminer le substrat des circuits de réentrée, la cause principale de l'AV liée à la cicatrice. Néanmoins, il s'agit d'une procédure longue et à haut risque, dont le principal problème reste la localisation imprécise du substrat.

En utilisant des méthodes Deep Learning (DL), nous avons étudié l'hétérogénéité de la cicatrice myocardique post-infarctus dans le ventricule gauche (VG) et sa relation avec l'AV, avec des données rétrospectives de 600 images de scanner cardiaque. Dans le contexte de la localisation du substrat et du guidage de la AR, il a été démontré que la paroi du VG plus mince dans la zone de la cicatrice a des propriétés de conduction lente. Notre objectif principal est de construire un modèle de prédiction pour analyser l'épaisseur du VG, améliorant ainsi la sélection des patients avant le traitement par AR ou une stratégie de prévention. L'objectif secondaire est d'explorer l'explicabilité du modèle. L'idée est d'identifier les régions spécifiques qui ont le plus d'influence sur la prédiction du modèle, ce qui, dans le contexte de la prédiction de l'AV, pourrait se traduire par une identification personnalisée des régions arythmogènes.

Nous avons développer un pipeline automatique pour transformer l'entrée (les scanners cardiaques 3D) en une carte polaire de la valeur de l'épaisseur du VG. Pour le modèle de prédiction, nous avons construit le modèle conditional variational autoencoder classification (CVAE-Class). Pour étudier l'explicabilité du modèle, nous avons utilisé la méthode GradCAM++. Nous avons montré que le modèle CVAE-Class était capable d'obtenir des prédictions AV plus précises par rapport aux marqueurs cliniques, y compris

la FEVG. La carte d'activation générée avec GradCAM++ a montré une forte corrélation entre les régions à fort coefficient et les régions amincies, ce qui a renforcé la relation entre l'amincissement de la paroi du ventricule gauche et l'arythmogenèse. Nous avons également étudié la méthode graph neural network, qui s'est avérée plus performante que le modèle CVAE-Class. Nous avons exploré l'explicabilité du modèle graphique à l'aide de deux méthodes : Local Interpretable Model-agnostic Explanation (LIME) et les integrated gradients, qui produisent des régions de coefficient élevé plus distinctes par rapport à la méthode GradCAM++.

Ce travail de recherche fait progresser nos connaissances en proposant une nouvelle méthode automatique pour analyser l'hétérogénéité de l'amincissement du VG, qui nous a permis de prédire de manière plus robuste le risque d'AV lié à la cicatrice, par rapport aux marqueurs cliniques actuels. En particulier, l'inclusion de l'explicabilité dans la sortie conduit à une meilleure transparence du modèle de prédiction. De plus, l'amélioration ultérieure de notre méthode, offre une perspective optimiste pour la traduction de ce travail dans la localisation automatique des substrats arythmogènes du AR.

Mots-clés: Arythmie cardiaque; Intelligence Artificielle; Modèle Explicable; Imagerie cardiaque; Multi-modalité.

Abstract

Sudden cardiac death (SCD) is a major health concern responsible for up to a yearly mortality of 400,000 in the US, and a similar figure in Europe. SCD is a consequence of the sudden and total stop of cardiac activity, also known as the sudden cardiac arrest (SCA). A study monitoring the cardiac activity has linked the SCA to ventricular arrhythmia (VA), which is responsible for up to 70% of all recorded SCAs. The preventive approach to VA and the patient risk of the SCD is the implantable cardioverter defibrillator (ICD), while catheter-based radiofrequency ablation (RFA) is a potentially curative therapy. ICD continuously monitors the cardiac electrical activity and is programmed to deliver appropriate electrical shock(s) in order to stop any detected VA episodes. However, a main limitation is the patient selection criteria, often based only on the left ventricular ejection fraction (LVEF) index, which suffers from both large number of false positive and negative identifications. RFA is an interventional procedure to eliminate the substrate of reentry circuits, the primary cause in scar-related VA, which can effectively prevent future episode of VA. Nonetheless, this procedure is a very time consuming, high-risk and error-prone procedure, for which the major issue remains the inaccurate substrate localization.

Based on novel Deep Learning (DL) methods, we studied the heterogeneity of the post-infarct myocardial scar in the left ventricle (LV) and its relation to VA, using a large retrospective dataset of 600 cardiac CT images. In the context of substrate localization and RFA guidance, thinner LV wall within the scar zone has been shown to have slow conduction properties. Our primary goal is to build a DL prediction model to analyse the LV wall thickness, thus improving the patients selection prior to the RFA treatment or a prevention strategy. The secondary goal is to explore the DL prediction model explainability. The idea is to pinpoint the specific regions on the model input that are most influential to the model prediction, which in the context of VA prediction could be translated into a personalised identification of arrhythmogenic regions.

We started by implementing a fully automatic pipeline to transform the input (i.e., 3D CT cardiac images) into a 2D bullseye representation of the LV thickness value. For the VA prediction model, we built a conditional variational autoencoder classification (CVAE-Class) model. To study the explainability of the model, we employed the GradCAM++ method. We successfully showed that CVAE-Class model was able to achieve higher accuracy VA predictions compared to the available clinical markers, including the LVEF.

The class activation map generated with GradCAM++ showed a high correlation of the high coefficient regions with the thinning regions, which solidified the relation between of LV wall thinning and arrhythmogenesis. We further explored the graph neural network method, which is shown to achieved better performance than the CVAE-Class model. We explored the explainability of the graph model using two methods: Local Interpretable Model-agnostic Explanation (LIME) and integrated gradients, which output smaller and more distinctive regions of high coefficient compared to the GradCAM++ method.

Overall, this research work advances our knowledge by proposing a novel fully-automatic method to analyse the heterogeneity of the LV wall thinning, which allowed us to predict more robustly scar-related VA risk, compared to the current clinical markers. In particular, the inclusion of explainability in the output is a critical feature leading to a better transparency to the prediction model. Moreover, the subsequent improvement of our method via integration of LIME and integrated gradient, provides an optimistic perspective for the translation of this work into the clinical routine for automatic localisation of the arrhythmogenic substrates of ablation therapy.

Keywords: Cardiac Arrhythmia; Artificial Intelligence; Explainable Learning; Cardiac imaging; Multi-modality.

Acknowledgements

First and foremost, I would like to thank my two thesis directors Maxime Sermesant and Hubert Cochet. Thank you for responding to my internship application and for backing me up for the audition of this PhD programme. The work of this thesis was a direct result from your supervision, support and encouragement during all these years. I would like to thank the members of the jury, Professor Elio Marijon and Ender Konukoglu for their precious helps reviewing this manuscript and their participation in my thesis defence. Finally, I would like to express my gratitude to Professor Pierre Jais for have accepted to chair my thesis jury.

Thanks to our collaboration centres and all the collaborators at the CHU Bordeaux and the Sunnybrook Research Institute (Dr. Mihaela Pop lab) for graciously sharing the data with us and for all the hard works in the data curation.

Thanks to my colleagues and collaborators, Marta Nuñez-Garcia, Nicolas Cedilnik, Tania Bacoyannis and Sonny Finsterbach for their helps in the realisation of this thesis. I would like to dedicate my special thank to Dr. Mihaela Pop, who has helped me tremendously with the writing of the manuscript.

Thanks to Prof. Nicholas Ayache, Prof. Hervé Delingette, Prof. Xavier Pennec, Prof. Marco Lorenzi and Dr. Irene Balelli, as well as all the members of the Epione team for making the working environment so dynamic and stimulating. I would not be so ungrateful to not mention Isabelle Strobant, who has helped me through all these years with the administrative tasks. I would also like to extend my gratitude to the team in the IHU Liryc and the CHU Bordeaux, who have contributed to the realisation of the works in this thesis.

Thanks to my parents in Cambodia for their supports and for always prioritise my education and without whom I would not be writing this manuscript. Last but not least, thanks to the three cats, Réglisse, Gessie and Gatsby and their owner, who have kept me accompanied and sane through the COVID-19 confinement and through this PhD programme.

Financial Support

The research leading to these results has been supported by the French Government through UCA Doctoral School EDSTIC, the National Research Agency (ANR) 3IA Côte d'Azur (ANR-19-P3IA-0002), and IHU Liryc (ANR- 10-IAHU-04). The project was also supported by the Inria Sophia Antipolis - Méditerranée, "NEF" computation cluster.

Contents

Glossary:

ACVD: Approximated Centroidal Voronoi Diagrams.

AHA: American Heart Association.

AI: Artificial Intelligence.

AUC: Area Under Curve.

CMR: Cardiac Magnetic Resonance.

CNN: Convolutional Neural network.

CT: Computed Tomography.

CVAE: Conditional Variational Autoencoder.

CVD: Cardiovascular Diseases.

DL: Deep Learning.

ECG: Electrocardiogram.

EF: Ejection Fraction.

EHRA: European Heart Rhythm Association.

FCN: Fully Connnected Network.

GCN: Graph Convolutional Network.

GradCAM++: Gradient Class Activation Mapping ++.

ICD: Implatable Cardioverter defibrillator.

KD trees NN: K-dimensional Trees Nearest Neighbour (search).

LGE: Late Gadolinium Enhancement.

LIME: Local Interpretable Model-agnostic Explanation.

LV: Left Ventricular.

LVEF: Left Ventricular Ejection Fraction.

MI: Myocardial Infarction.

ML: Machine Learning.

MRI: Magnetic Resonance Imaging.

MYO: Myocardium.

RFA: Radiofrequency Ablation.

ROC: Reveiver Operator Curve.

RV: Right Ventricular.

SAX: Short-Axis view.

SCA: Sudden Cardiac Arrest.

SCD: Sudden Cardiac Death.

SDA: Style Data Augmentation.

VA: Ventricular Arrhythmia.

VAE: Variational Autoencoder.

VF: Ventricular Fibrillation.

VT: Ventricular Tachycardia.

WT: Wall thickness.

1 Clinical and Methodological context and background

Contents

The heart has always been regarded as the most vital organ in the human body across different civilisations, cultures and religions. Thanks to modern biology, medicine and imaging-based investigative technologies, the intricate structure-function interaction of the heart has been better understood in both normal and pathological cases. Every heartbeat is an orchestrated electrophysiological process, and a major disturbance in the heart rhythm could lead to catastrophic events, including death. Therefore, finding the location of potential arrhythmogenic sources is a primer objective in identifying the potential patients at risk of lethal events.

1.1 Introduction

1.1.1 Motivation

Cardiovascular disease (CVD) is the leading cause of death worldwide, accounting for approximately 18 million deaths per year according to the World Health Organization. While this term can be refer to multiple heart (cardio-) and vessel (vascular) conditions, a critical problem related to CVD stems from the consequences of inadequate blood flow to the organs of the body. Each cell requires a continuous exchange of gases,

fluid, electrolytes and macromolecules, all of which depend on the circulatory system. Therefore, the failure of the cardiovascular system to adequately pump blood into the various organs of the human body via the circulatory apparatus, may lead to serious and even deadly events. The most important such organ is the brain in the nervous system, for which the lack of oxygenated blood is a primary cause of death in CVDs.

Sudden cardiac death (SCD), is an unexpected death that occurs due to sudden cardiac arrest (SCA), the latter being characterised by a malfunctioning of the heart's electrical system which causes irregular often abnormally high rhythms. The sudden cessation of blood flow to the brain leads to an immediate loss of consciousness, often impeding the patient ability to seek external help, and leading to a small survival rate of 10% [Al-Khatib, 2018]. One of the major causes of SCD is ventricular arrhythmia (VA). Ventricular tachycardia (VT) is a cardiac rhythm disorder defined by an abnormally fast heart rhythm that can degenerate into a disorganised and chaotic propagation of the electrical wave, a condition known as ventricular fibrillation (VF). A study monitoring the patients leading up to the sudden cardiac arrest (SCA), showed that up to 70% of the SCA cases have been direct consequences of VA episodes. Thus, patient-specific predictions of SCD/SCA risk caused by VA is of paramount importance. With this respect, modelling based on imaging data could offer a powerful non-invasive method to predict VA risk.

The work done throughout this thesis is focused on leveraging a data-driven approach to build an automatic model for accurate VA predictions from cardiac 3D images. Specifically, here we investigate several automatic image processing and deep learning methods in order to develop a fully-automatic pipeline for VA risk prediction from imaging data.

The focus was also put on the explainability and interpretability of the prediction model. Several methods were used to generate an explainability coefficient map, which can highlight the region most "influential" for the model decision. The primary use of this explainability study was to allow some transparency of the deep learning model, thus increasing our confidence towards the model. Furthermore, it also allows a better understanding and identification of the specific arrhythmogenic site in the input. Our original pipeline can be rapidly translated into the clinical routine investigations, bringing us one step closer to an automatic detection of the VA substrates (i.e., so called reentry channels), which are the target of radio-frequency ablation, a potential curative catheter-based intervention for VA.

1.1.2 Manuscript Organisation & Contributions

The first chapter of the manuscript covers the motivation, as well as the clinical and technical context of this thesis. For the clinical context, we provide a brief introduction of the cardiac anatomy, physiology and pathology. Next, a clinical background describing

VA caused by myocardial infarction (MI) is included, along with the most commonly used scar imaging techniques such as MRI and CT imaging, for MI diagnosis. For the technical context, the basics of deep learning and explainable learning are provided. Notably, more context for each network architecture will be provided in the corresponding chapters, since several model architectures were used in this thesis.

In chapter 2, we describe in detail our novel work on automatic segmentation of MR images. This chapter is split into two parts: the multi-modality and mono-modality segmentation, respectively, In the first part, we proposed a data augmentation method to integrate images from other MR imaging methods that were used to compensate for the limited number of the available late gadolinium enhanced (LGE) MR images of infarct scars. We showed that the method was able to increase the segmentation accuracy, allowing us to perform the fitting of a generalised segmentation model capable of segmenting several MR modalities (as input) at the same time. In the second part, we introduced a direct anatomical constrain from the relation between different structures of the LV: the epicardium, the endocardium; and, the myocardial scar. Although this method was proven to improve the previous scar segmentation accuracy, the accuracy was still considered low for any concrete analysis of the myocardial scar.

In the light of these limitations, we shifted our attention in the following chapters to the analysis of cardiac CT images of post-infarction patients, for which a large clinical dataset was made available through our collaboration with the University Hospital of Bordeaux. In chapter 3, we proposed a fully automatic pipeline to derive the LV wall thickness and to generate the 2D bullseye representation from cardiac CT images of the above mentioned database. This representation allows our model to specifically focus on heterogeneous regions of the scar (i.e., where VA substrate may reside) that can be characterised by a thinning of the left ventricle (LV), while at the same time removing the shape discrepancies between the different LV inputs.

In chapter 4 and chapter 5, we designed two different DL methods for VA prediction. In chapter 4, we proposed a conditional convolutional variational autoencoder classifier model (CVAE-Class) using convolutional layers on a 2D bullseye input. We then applied an explainability study, using a gradient back-propagation method (GradCAM++), which allowed us to better understand the region (in the input) that was most influential to the model decision.

In chapter 5, as an improvement to the previous method, we further proposed a graph DL model using a graph convolutional network (GCN) layer. We also proposed new methods such as LIME and integrated gradients, to study the interpretability of our model. These advanced methods generate more specific arrhythmogenic regions, at least as viewed by the trained model.

Finally, the summarising conclusion of this work is presented in chapter 6, where we detailed our original contributions to the field, specifically regarding the application of DL methods for automatic detection of VA associated arrhythmogenic regions. We also briefly included some prospective ideas for potential future works.

1.2 Clinical Context

To properly understand the work in this thesis, the following sections of this chapter will be dedicated to provide the clinical (section 1.2) and technical (section 1.3) contexts related to our work. Figure 1.1 shows a schematic diagram highlighting the fundamental contexts of our work.

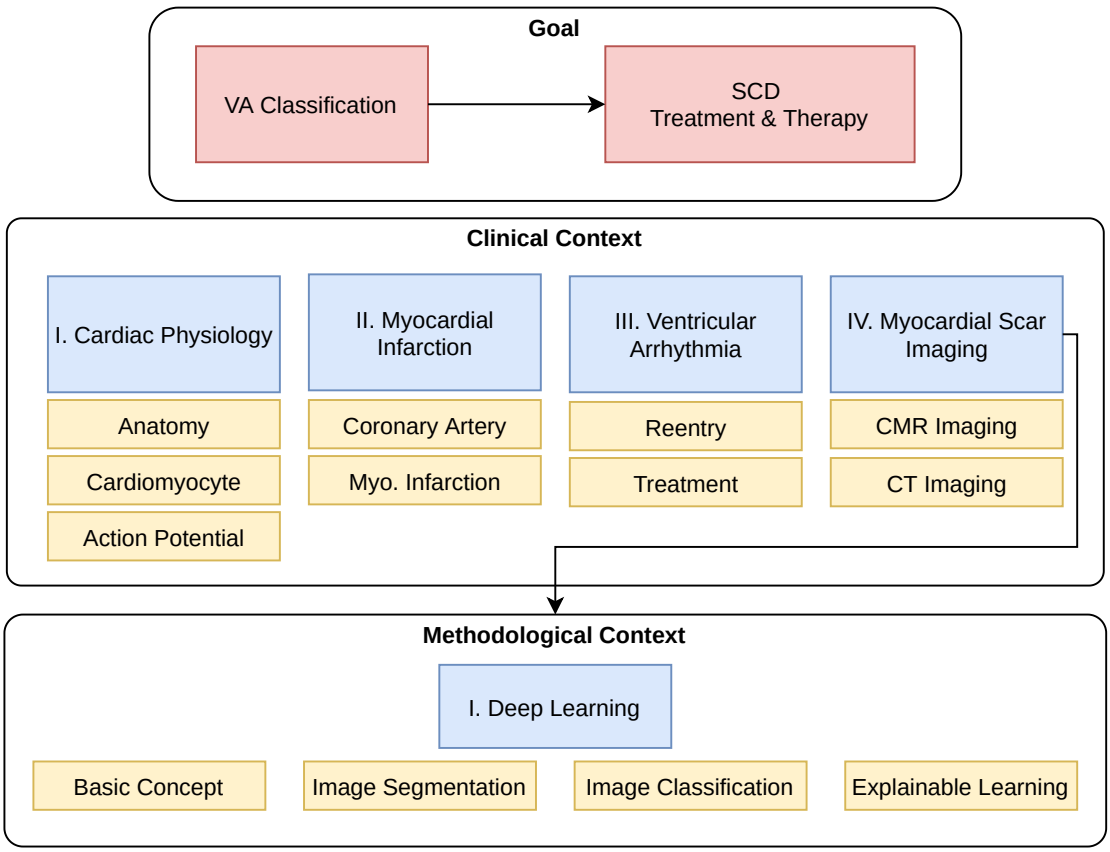


Fig. 1.1.: Diagram of the clinical and methodological context in relation to the main goal of this thesis. For the clinical context, the cardiac physiology is detailed, which serves as the basis for the understanding of pathology of myocardial infarction and scar-related VA. Note that myocardial scar imaging is the main entry point into the methodology of this work. For the methodological context, we provide details concerning: deep learning including its basic concept; its specific applications used in this work; and, explainable learning.

1.2.1 Cardiac Physiology

1.2.1.1 The Anatomy of the Cardiovascular System

As shown in Figure 1.2, the cardiovascular system, also known as the circulatory system, is comprised of the heart, blood and blood vessels. The lymphatic system serves as a complementary system. The main function of the circulatory system is to deliver the ions and molecules (primarily in the form of nutrients and oxygen-rich blood), to different parts of the human body. The heart is a muscular organ and its primarily function is to pump the blood. Briefly, during the heart contraction, the blood is pumped out off two ventricles through some major vessels: the deoxygenated blood flows from the right ventricle (RV) through pulmonary into the lungs through pulmonary artery, while the oxygenated blood flows from the left ventricle (LV) through the aorta into the systemic circulation, delivering oxygen and nutrients to all organs. The deoxygenated blood returns back to the heart via the venous system and enters the heart through two atrial chambers (right and left, respectively).

The heart is divided into four chambers: left and right ventricles, and left and right atria. The two atria are separated by the interatrial septum, and the ventricles by the interventricular septum. These walls are closed in order to prevent the blood to flow between the chambers. Between the right atrium and right ventricle is the tricuspid valve or the right atrioventricular valve, and between the left ventricular and left atrium is the mitral valve or the left atrioventricular valve. The two other valves playing an important role in the cardiac cycle are the aortic valve (located between the aorta and the left ventricle) and the pulmonary valve (located between the pulmonary arteries and the right ventricle).

The valves are periodically open and closed, allowing the blood to flow between the chambers and contributing to the regulation of pressure during the different phases of cardiac cycle.

There are 2 general phases in the cardiac cycle: systole and diastole. The systole is marked by the contraction and ejection of the ventricles, during which the AV valves are closed while the aortic and pulmonary valve are open. The diastole corresponds to the ventricular relaxation and filling phase. During this phase, the aortic and pulmonary valves are closed, while the mitral and tricuspid valves are open, thus allowing the blood to flow freely into the ventricles. Towards the end of the diastole the atria contract further, pumping the blood into the ventricles and increasing the ventricular pressure before the next systole phase starts.

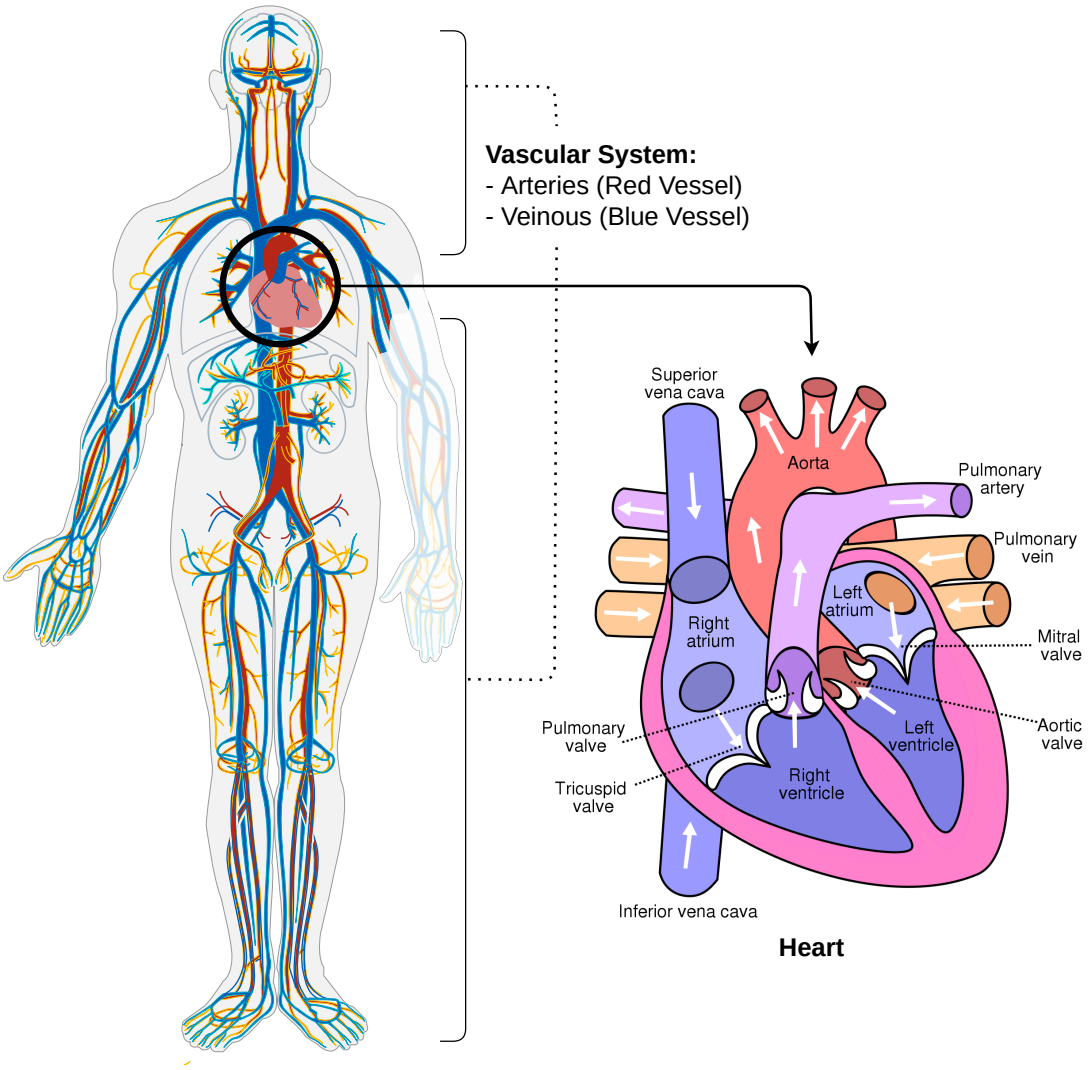


Fig. 1.2.: The global overview of the cardiovascular system. The heart and vessels ensure the mechanical circulation of the blood through different organs of the body. The white arrows denote the direction of the blood circulation within the heart through the 4 chambers and valves, during the cardiac cycle. (illustrations by LadyofHats from Wikimedia commons, Public Domain, and Wapcaplet from Wikimedia commons, CC-BY-SA 3.0 license.)

1.2.1.2 Cardiomyocyte

At the cellular level, the heart muscle is comprised of cardiomyocytes (30%) and non-cardiomyocytes (70%), the majority of which are cardiac fibroblasts [Jugdutt, 2003]. Although they constitute the majority of cardiac cells, the fibroblasts do not play an active role in the cardiac contraction function. Instead, they play an important role in cardiomyocyte proliferation and extracellular matrix homeostasis (i.e., degradation and synthesis) [Porter, 2009]. The extracellular matrix is a protein network comprised in majority of collagen, which maintains the structural integrity of the heart.

Figure 1.3 illustrates the structural composition of the cardiomyocyte. The cardiomyocyte is a type of striated muscle cell, named so because of the cross striations formed by the alternating dark (A) bands and light (I) bands, as observed under the microscope. A similar structure has been observed in the skeletal muscle. The striation is the marking of the contractile unit of the myocyte, called the **sarcomere**. The sarcomere is identified as the segment between two Z-lines, which is the transverse line in the middle of the I band. The dark color observed in the A band corresponds to the presence of the myosin, a large molecular weight protein. A bundle of up to 300 myosins creates a **thick filament**, giving the entire section a dense dark color. The myosin filament is connected to the Z-line via a protein **tintin**, which keeps the thick filament contained within the sarcomere. In the I band, we can partially observe the **thin filament**, composed of a protein **actin**, The thin filaments are arranged hexagonally around the thick filament. It is the interaction between the actin and myosin that leads to muscle contraction, as the actin and myosin filaments slide past each other, shortening the sarcomere length.

However, in contrast to the skeletal muscle cells which are multinucleated and run along the entire muscle, the cardiomyocytes are single-nucleus cells interconnected into a **functional synctyium**. The functional syncytium is defined as a group of cells that function as a single unit, a feature possible for the cardiomyocytes as they are connected at the **intercalated disks**. The intercalated disks are composed of **gap junctions**, which are proteins channels that allow the flow of ions from one myocardial cell to another, thus propagating the electrical impulse between interconnected myocytes. Once electrically excited, the myocyte contracts. The syncytium allows a synchronised contraction of the heart muscle.

1.2.1.3 Cardiac electrophysiology: Action Potential

The change in the electrical property of the muscle cell triggers the interaction between the proteins inside the cardiomyocyte, which in turn triggers the contraction of individual cells. In its resting state, the membrane potential of a cardiomyocyte is maintained at -90mV by a balance of the positively charged ions (Na+, K+ and Ca+) and the negatively charged proteins. Compared to the interstitial concentration, the K+ ions have an outward gradient (i.e., higher concentration inside the cell), while for the Na+ and Ca+ ions the gradients are inward oriented. The exchange of all these ions between the cell and the interstitial space is done via multiple ion-specific channels. There are two general types of ion channels: voltage-gated and receptor-gated. The voltage-gated channels open and close as a reaction to the change in membrane potential, and are the main channels involved in action potential. In contrast, the receptor-gated channels react to the presence of chemical signals released by other regulators of the heart, such as acetylcholine (a neurotransmitter released by vagus nerves) which causes a reaction to potassium channels.

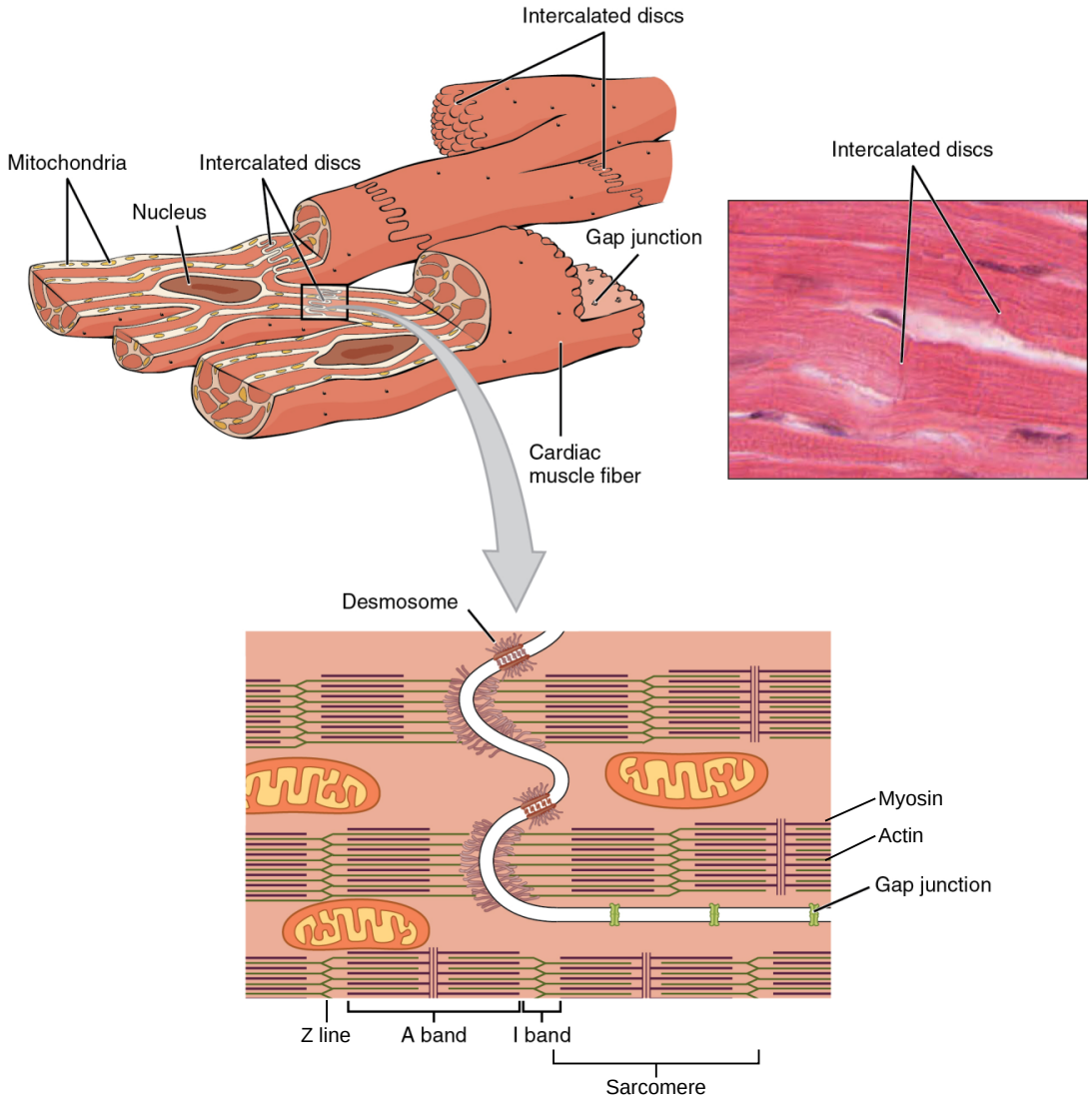


Fig. 1.3.: The structural composition of the cardiomyocyte. The cardiomyocyte is a mononucleated cell, which is one of the main difference from skeletal muscle cell. The cardiomyocytes are interconnected at the intercalated discs, which create the functional syncytium. The gap junctions located at the intercalated discs, facilitate the propagation of the electrical impulse between myocytes. The cellular composition consists of the thick (myosin) and thin (actin) filaments, with dark (A) and light (I) bands, similarly observed in the skeletal muscle. (illustration by Open Stax Anatomy & Physiology, CC-By license.)

The action potential represents a change in transmembrane voltage, typically starting with a rapid electric excitation of the membrane named depolarisation phase, which is followed by a "slower" relaxation back to the resting state of the voltage (i.e, the repolarisation phase). The action potential triggers the contraction process, which is then followed by cardiomyocyte's relaxation. The depolarisation is called so because of the flux of positively charged ions into the myocyte, which reduces the negative polarity of the membrane potential from its resting state (-90 mV). The repolarisation corresponds to to a phase when the ionic balance within the myocyte is re-established

through internal ionic pumps, negatively charging the membrane back to its resting state. There are two types of action potentials, as defined by the source of trigger: pacemaker and non-pacemaker action potentials, respectively.

The pacemaker action potential undergoes a depolarisation triggered by the specialised cells generating spontaneous action potentials. This depolarisation is achieved by a "relatively" slow influx of Ca+, which gives the name "slow response". The primary cardiac pacemaker is located in the **sinoatrial node**, which can be found in posterior wall of the right atrium (RA). The secondary pacemaker cells can be found along the cardiac conduction system, although their pacemaker activity is suppressed by the **overdrive suppression** from the primary pacemaker and could reactivate when the electrical activity from the primary pacemaker is inhibited. The normal cardiac rhythm is controlled by the SA node; therefore, it is also called **sinus rhythm**.

The non-pacemaker action potential is the depolarisation wave propagated from the adjacent cells. This type of action potential is what excites the majority of the myocytes in the heart, and the phenomenon is also referred to as cell-to-cell conduction. After the first cell is depolarised, the positive charge ions accumulate and flow towards the adjacent cells through the gap junctions of low electrical resistance (i.e., high conductivity), charging those. When these adjacent cells are sufficiently depolarised (i.e., the voltage reaches a certain threshold), an action potential is triggered. The propagation then continues in a similar fashion to the following neighbouring cells. Because of its rapid depolarisation, the non-pacemaker action potential is referred to as "fast response". The non-pacemaker action potentials can be divided into 5 phases. The phase 0 is defined by the rapid depolarisation followed by the initial repolarisation in phase 1. The phase 2 is marked by a delayed repolarisation, where the membrane potential has a plateau. A rapid repolarisation occurs in phase 3, while the membrane potential reaches back its resting state in phase 4.

The refractory period refers to the duration between phases 0-3, when the change in the membrane potential does not lead to a depolarisation. It is only during the phase 4 when the next depolarisation wave could be re-triggered. This is a preventative measure against a constant re-excitation by a single depolarisation wave. Nonetheless, abnormal action potential generation at the cellular level, such as the abnormal automaticity and triggered activity (spontaneous depolarisation in phase 3 to early phase 4), can lead to self-sustaining action potentials and tachycardia. Post-MI arrhythmia can also be generated at the structural level by a mechanism called anatomic reentry, which is caused by abnormal conduction pathways formed by surviving bundles of slow conduction that traverse dense scars.

Figure 1.4 details the pacemaker and non-pacemaker action potential of the different components of the heart.

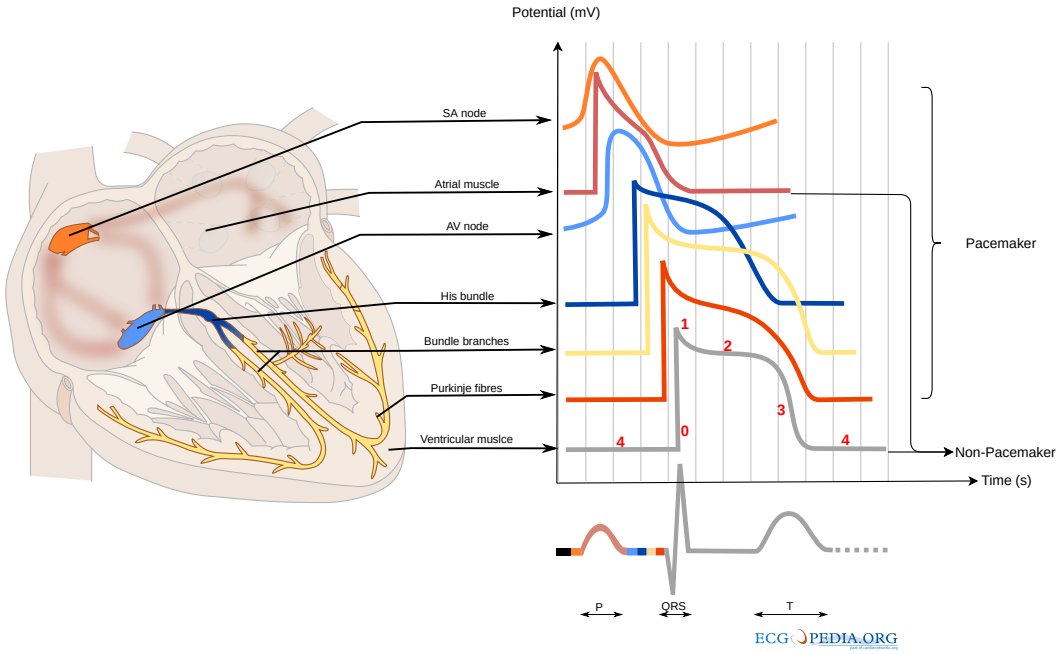


Fig. 1.4.: Different cardiac action potential. The pacemaker action potential can be automatically generated by the specialised cells along the conduction system (SA node, AV node, His bundle, bundle branches and Purkinje fibres). The traceable ECG signal results from the summation of all action potentials propagating within the heart. (illustration by ecgpedia from Wikimedia commons (CC BY-SA 3.0 license))

1.2.2 Myocardial Infarction

1.2.2.1 Coronary Artery: Cardiac irrigation system

The heart is a continuously beating aerobic muscle, and as such, it requires an elaborated circulation system to supply the oxygen needed for a high demanding metabolism. Two main coronary arteries (left and right) are responsible for this task, both starting at the root of the ascending aorta, next to the aortic valve. The left main coronary artery splits into the left anterior descending artery (along the anterior surface of the interventricular septum), and the circumflex artery which is located on the posterior side at the intersection between left ventricle and right ventricle. The left main coronary artery and its branches supply the blood to both left atrium and left ventricle. The right main coronary artery descends along the right atrium and right ventricle, branching out to irrigate the right chambers of the heart.

The capacity of the healthy coronary arteries exceed the capacity of the myocardial micro-circulation, maintaining an over-supply of oxygenated blood to the heart. As such, a certain degree of obstruction in the coronary arteries can be tolerated, and a certain degree of coronary atherosclerosis can be ubiquitous and tolerated among middle age to

older population. The arteriosclerosis is an inflammatory disease defined by a cumulative and harden plaque, stiffening and narrowing the vessel lumen. [Gopalan, 2022] The reduced size of lumen may be tolerated up to 75% of its normal size, until there is an impact to the myocardium demand at the period of exertion [Steenbergen, 2012]. It is only when the blood supply does no longer meet the demand, that the ischaemia and its symptoms would arise, in which case the main symptoms of hypoxia (i.e., lack of oxygen) is the chest pain or angina pectoris. Moreover, the coronary system can be resilient as the progressive narrowing of the vessel lumen leads to so-called collateralisation, a development and enlargement of nearby vessels that can help supply blood to the ischaemic regions. Nonetheless, the sudden and complete block of the blood supply can lead to a condition called myocardial infarction or heart attack. Figure 1.5 shows the branches of the coronary arteries and the partially blocked blood vessel caused by coronary atherosclerosis.

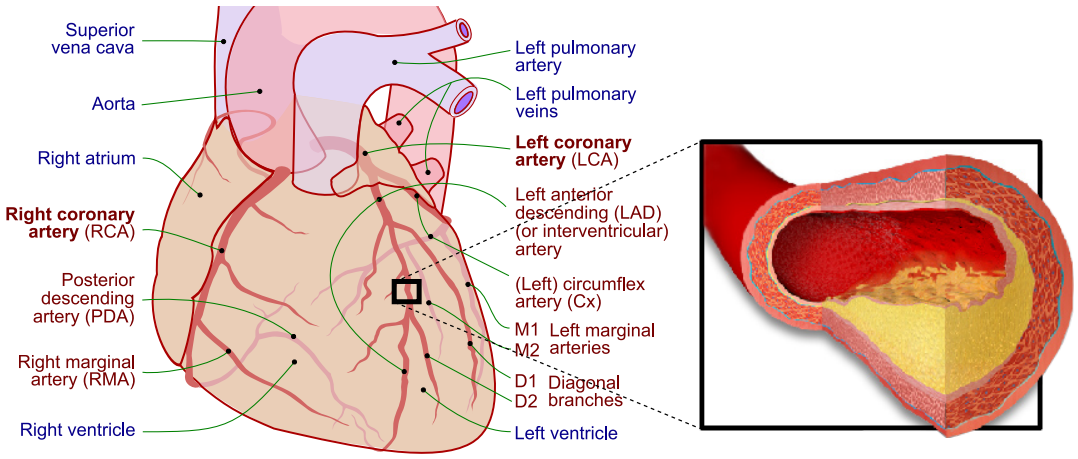


Fig. 1.5.: Coronary branches and arteriosclerosis. Sudden and complete blockage at any point in the coronary branch would lead to the infarction of the downstream territories. (Part of illustration by Patrick J. Lynch from Wikimedia commons, CC BY-SA 3.0 license, and BruceBlaus from Wikimedia commons, CC BY-SA 4.0 license)

1.2.2.2 Myocardial Infarction

Myocardial infarction (MI) is the result of a complete loss of blood supply to a myocardial region. MI is caused by an acute occlusion of a major coronary artery, which can be due to thrombus formation and the vasospasm subsequent to the rupture of an arteriosclerosis plaque. Within the first ten seconds following the blood supply cut, the anorexic metabolism starts leading to the concentration of noxious products (e.g. lactic acid). After 1 minute, hypo-contractility occurs in cardiomyocytes. Then, within the next few minutes, reversible structural changes can be observed such as cellular and mitochondrial swelling, and distortion of the transverse tubular system. Irreversible changes followed by death occur after a prolonged state of sustained ischaemia in cardiomyocytes.

Several studies have demonstrated a heterogeneity in myocardial necrosis onset, starting from the subendocardial region (under the lumen of the heart chamber), the midmyocardium and the subepicardium region, respectively [Steenbergen, 2012]. The theory behind this phenomenon is based on a higher demand and limited supply of oxygen in the insulted region closer to the endocardium [De Tata, 1983; Hoffman, 1985; Boatwright, 1980; BUCKBERG, 1972; Spaan, 1981; Downey, 1975]. As such, the process of necrosis can start in a small subendocardial cardiomyocyte as early as the 15-20min after the coronary occlusion. As the occlusion persists, the necrotic area then expands to less vulnerable areas. Therefore, the size and transmurality of the infarct region greatly reflects the severity of the myocardial infarction.

The healing process follows the local necrosis phase. During healing, a structural and functional remodelling phase takes place, starting with a sudden inflammatory reaction within the infarcted area and ending with the formation of collagen-based scar [Frangogiannis, 2008]. This process can be divided into three phases: inflammatory phase; proliferative phase; and, maturation phase [Frangogiannis, 2006]. The rupture of a cell membrane results in a redistribution of the intracelluar contents into the interstitial space, which triggers an inflammatory reaction in order to clear all the dead cells and extracellular matrix debris. The proliferative phase or the reparation phase is marked by the activation of the cardiac fibroblasts, which triggers the myofibroblasts transdifferentiation (i.e., the transformation from cardiac fibroblasts to myofibroblasts), along with the proliferation and synthesis of extracellular matrix. To accommodate the high demand for oxygen and nutrients during this phase, an extensive micro-vascular network is being formed throughout the healing region. The healing reaches its final phase after the apoptosis of the myofibroblasts and the maturation of vascular cells, its final product being a cross-linked collagen-based scar. As they mature, some vessels also undergo apoptosis, whereas some vessels are being coated by mural cells to prevent regression, which results in the retention of a low capillary density within the scar region.

At the end of the healing period, the infarcted tissue cannot reverse to a normal structure or normal cellular composition. This process results in a compensating structural response of the ventricular tissue referred to as "ventricular remodelling". The temporal and local remodelling includes the thinning of infarcted myocardial wall and a hypertrophy of the non-infarcted area, as well as a global dilation and an increased sphericity of the respective heart chamber. Moreover, the structural changes and geometrical configuration of the scar serve as the base of anatomic reentry mechanism in ventricular arrhythmia. Figure 1.6 illustrates a histological comparison between healthy myocardium and dense fibrotic scar, post-infarction, as seen under light microscopy.

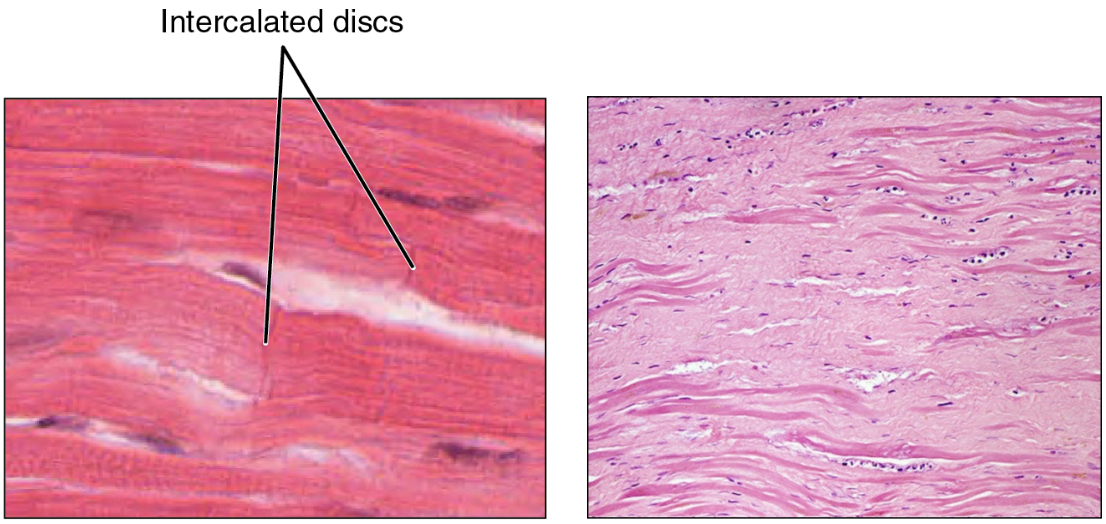


Fig. 1.6.: Comparison between healthy myocardium and dense scar zone. Note the surviving cardiomyocytes within the scar region, which are the base of reentry mechanism of ventricular tachycardia. (parts of the illustration was done by Open Stax Anatomy & Physiology, CC-By license, and Katarzyna Michaud et al from Wikimedia commons, CC-BY-SA 3.0 license.)

1.2.3 Ventricular Arrhythmia

Ventricular arrhythmia (VA) is an umbrella term covering all rapid, cyclical and abnormal heart beats occurring in the ventricle. The majority of VAs are in the form of ventricular tachycardia (VT), which can degenerate into ventricular fibrillation (VF). The classification of VAs is based on the heart beat rate and the QRS morphology. The VT and its classes have a rate of higher than 100 beats/min, while VF is defined as a rapid and chaotic rhythm of more than 300 beats/minute.

As mentioned in section 1.2.1.3, some causes of VA can be at the cellular level, such as the abnormal automaticity and trigger activity. However, the understanding and early diagnosis of these mechanisms are still somewhat limited. On the other hand, recent studies have made great strides in understanding the VA mechanism at the macroscopic level, especially regarding the mechanism of anatomic reentry related to structural heart disease, such as chronic MI [Anter, 2018; Ciaccio, 2022].

The current statistics indicate that the risk of VA related SCD is high within the first 6 months after the acute MI and only re-peaks after 4 to 10 year post-MI [Issa, 2018]. This delayed onset of deadly VA poses a great concern in SCD management and constitutes the fundamental interest of the work in this thesis.

1.2.3.1 Reentry Mechanism

The reentry mechanism involves a conduction block or a slow conduction zone and the surrounding fast conducting tissue. The structure of the healthy myocardium allows a fast cell-to-cell conduction of the electrical impulse, as explained in section 1.2.1.3. As the electrical wave propagates, the refractory period of the cardiomyocyte prevents re-depolarisation, until the next action potential wave arrives. However, the surviving cardiomyocytes and tissue remodelling within the scar region (see section 1.2.2), create a conduction block or a slow conduction zone where the action potential propagates at a much slower speed than in the healthy myocardium. The re-depolarisation is triggered when the action potential exits the temporary block, meeting the excitable cardiomyocytes in the healthy region. This favours a reentry circuit with a self-sustained action potential wave that will continue to propagate through the circuit. Figure 1.7 illustrates the difference between a normal action potential circuit and a reentry circuit, where slow conduction channel is present.

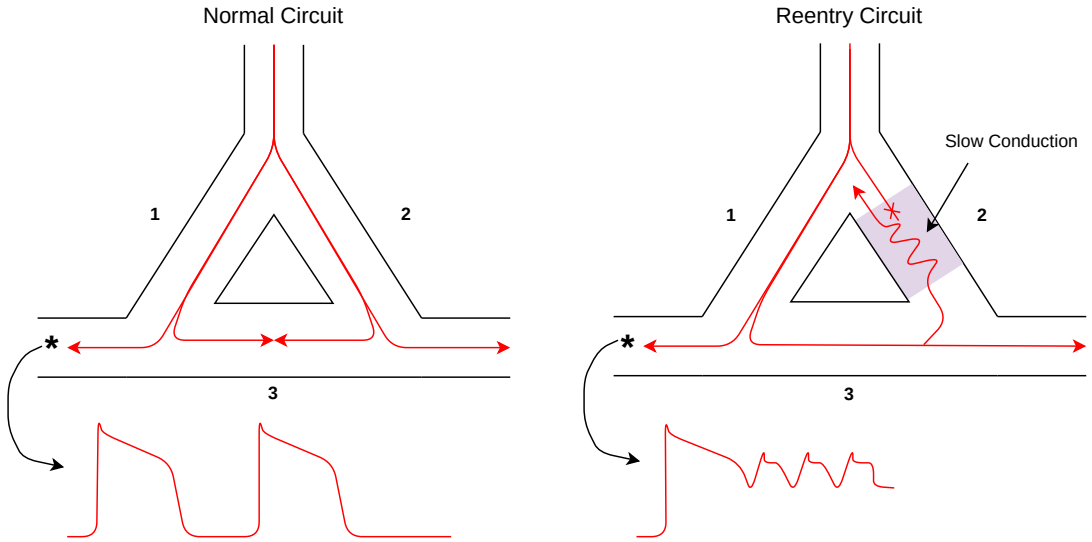


Fig. 1.7.: Reentry mechanism. In the normal circuit, the electric impulse travels down the branches 1, 2 and 3, while the refractory period prevents a re-triggering. In the reentry circuit, if the retrograde impulse passes through the slow conduction block and reaches the excitable cells (after the refractory period), a premature action potential is triggered down the branch 1. If this reoccurs with successive action potentials, then the arrhythmia occurs and is self-sustained. Illustration adapted from [Klabunde, 2011].

1.2.3.2 Treatment of VA

The first line of therapy to treat the VA is the antiarrhythmic drug therapy, which is prescribed to suppress the potential VT triggers. Nonetheless, the efficacy of this approach is still limited, with more than 40% of patients still experiencing sustained VT within the two years of therapy [Soejima, 2002]. Moreover, this approach can also result in undesirable side effects, including as pro-arrhythmic in patients with structural

heart disease. Therefore, under the current guideline, most anti-arrhythmic medications are advised to be administrated to patients already protected with the implantable cardioverter defibrillator (ICD).

Aside from medication, the clinical management of scar-related VA is also closely tied to the two available approaches: implantable cardioverter defibrillator (ICD) which does not treat VA but rather prevents SCD, and radio-frequency ablation (RFA) which is regarded as a potentially curative therapy.

The ICD is an implantable device that delivers anti-tachycardia pacing or a shock to the heart when a VA episode is detected, which resets the sinus rhythm. Therefore, the main benefit of ICD is to eliminate the VA episode and not to prevent its occurrence. However, ICD prevents sudden cardiac arrest and the potential lethal consequences. Following the current EHRA/AHA guidelines, the device is to be implanted preemptively to the patients at risk of lethal VA.

Under the recommendation for ICD patient selection as a primary prevention, the decision is mainly based on the value of ejection fraction of the left ventricle (LVEF), which is a clinical index that measures the relative change in LV volume between end-diastole and end-systole [Al-Khatib, 2018]. However, this strategy leads to nonessential implants, where up to 3/4 of the eligible patients would not receive any appropriate therapy within a 5-year post-implant period. Furthermore, it has been shown that based only on the recommended cutoff value for LVEF (i.e., at 35%) this type of selection misses more than 80% of the potential VA cases, where the SCD occurred in patients with preserved LVEF (i.e. LVEF > 35%) [Al-Khatib, 2018].The effectiveness of the LVEF index for VA prediction is further addressed in chapter 4.

In contrast, catheter-based RFA is able to precisely deliver radio-frequency energy that thermally ablates the arrhythmogenic substrate. Correct patient selection and accurate localization of ablation targets are critical in order to increase the success rate of the procedure, as well as to limit potential complications related to the intervention.

Through the complete elimination of the arrhythmogenic substrate, the recurrence of the arrhythmia can be prevented. The procedure takes place in the electrophysiology (EP) laboratory and typically employs an X-ray guided electro-anatomical mapping procedure to identify the VA substrate (i.e., critical channels part of the reentry circuits) to be ablated. To confirm the VA diagnosis, another invasive and time consuming procedure is performed, and involves a dangerous VA inducibility test [Al-Khatib, 2018]. Currently, the RFA success is limited by inaccurate and exhaustive identification of critical channels, by the inability to successfully induce the arrhythmia in some cases, and sometimes by inaccessible locations of arrhythmogenic substrate. Under the current guideline, the

ablation is proposed to patients who already suffered from multiple VA episodes and where the ICD therapy is in efficient or poorly tolerated.

The success of both approaches, ICD and RFA, relies heavily on the correct risk stratification of VA. However, for successful RFA interventions, the VA patient classification must be extended to include the localisation of specific arrhythmogenic sites relative to the scar location. This type of scar characterization is achieved in the clinics by using various cardiac imaging techniques, as described next.

1.2.4 Myocardial Scar Imaging

As discussed in section 1.2.3.1, the myocardial scar heterogeneity in the setting of chronic infarction, plays an important role in the mechanism of VA induction. The detection and quantification of the myocardial scar is therefore, the key in the diagnosis of scar-related VA. Non-invasive cardiac imaging techniques can be used to observe the scar tissue heterogeneities (i.e., dense scar vs surviving tissue harbouring the VA substrate). The most important techniques are cardiac magnetic resonance (CMR) and computed tomography (CT) [Berruezo, 2015]. Thanks to its wider accessibility and low cost, echocardiography is used more often (compared to CMR and CT) in the clinics for the evaluation of cardiac function and calculation of EF and other indices, [, 2022]. However, this technique can neither distinguish the scar from healthy tissue, nor identify the VA substrate. Thus, both CMR and CT imaging allow direct access to the topology of myocardial scar, making them more widely utilised for the planning of RFA interventions [Berruezo, 2015; Al-Khatib, 2018].

1.2.4.1 Cardiac Magnetic Resonance Imaging

CMR is an imaging technique based on an advanced technology that excites and then detects the change in the direction of the rotational axis of protons. These protons are found in the water that makes up living tissues, but can also belong to atoms of other compounds such as fat. The technique employs a powerful magnet which produces a strong magnetic field that forces protons in the body of the patient to align with that field. An RF current is pulsed through the patient stimulating the protons and moving them out of equilibrium. During the relaxation time (after turning the RF current off), the protons emit energy that is captured by special sensors (coils). Finally, the image is reconstructed based on different relaxation times of tissues, which offer different contrast between tissues.

Several MR imaging methods (or modalities) are used in CMR, depending on the type of evaluation required. For instance, CINE imaging acquires serial images of the beating heart during one heart cycle, allowing the clinicians to calculate the EF index from

segmented endocardial contours. In clinical practice, scout and cine MR images are usually obtained to define the short and long axis of the heart, as well as to assess the morphology and contractile function of the LV. Another method is late gadolinium enhancement (LGE) which uses gadolinium (Gd) as contrast agent, and is currently the gold standard for the evaluation of the myocardial scar [Al-Khatib, 2018]. The main procedure of LGE CMR involves the injection of a Gd-based contrast agent approximately 10min before the imaging starts, allowing a delayed enhancement (i.e., increased signal intensity) within the scar area to be visualized [Kim, 2003]. This delayed enhancement is a result of a slower washout rate of the contrast agent within the scar regions compared to the washout rate in the healthy myocardium. The total imaging time can be between 30-40mn [Kim, 2003].

Figure 1.8 shows an example of scar visualisation in the LGE images of a patients. The LGE images allow not only the direct scar visualisation but also its size quantification. Although CMR offers the best contrast among all imaging techniques, 2D LGE images are of low resolution (i.e., typical slice thickness being 8-10mm) and require manual segmentation by experienced clinicians to correctly delineate the scar region. On the technical side, it is important to note that MR imaging does not allow the imaging of patient with metal or magnetically-incompatible implants, due to dangerous interference with the strong magnetic field of the scanner.

1.2.4.2 Computer Tomography Imaging

The computer tomography (CT) is a modern imaging technique based on computerised X-ray imaging. The X-ray imaging employs an X-ray source emitting radiation through the patient, which is absorbed at different rate by different tissues according to the radiological density of the tissues. As the radiation passes through the patient, the remaining radiation is captured on the film or the X-ray detector, where the image is reconstructed. In contrast to the X-ray imaging, which only produces 2D images, the CT scan can be used to produce 3D images. This is done by employing a narrow beam of X-rays source and detector, which rotates around the patient and allows the reconstruction of 2D cross-sectional image slice at each full rotation. The image slice gives the name 'tomography'. The 2D slices are then be stacked together to produce a 3D image.

In cardiac imaging, the CT scan can be used to detect and characterise the underlying structural heart disease thanks to its excellent resolution, and under the current recommendation in the case where echocardiographic imaging is inadequate and CMR is contraindicated [Al-Khatib, 2018; Di Marco, 2017]. To characterise the myocardial scar region, several delayed enhancement techniques were also proposed for the cardiac CT [Esposito, 2016; Gerber, 2006; Lardo, 2006]. The techniques use iodine as the contrast agent, with the imaging being delayed between 5-10 minutes following the contrast

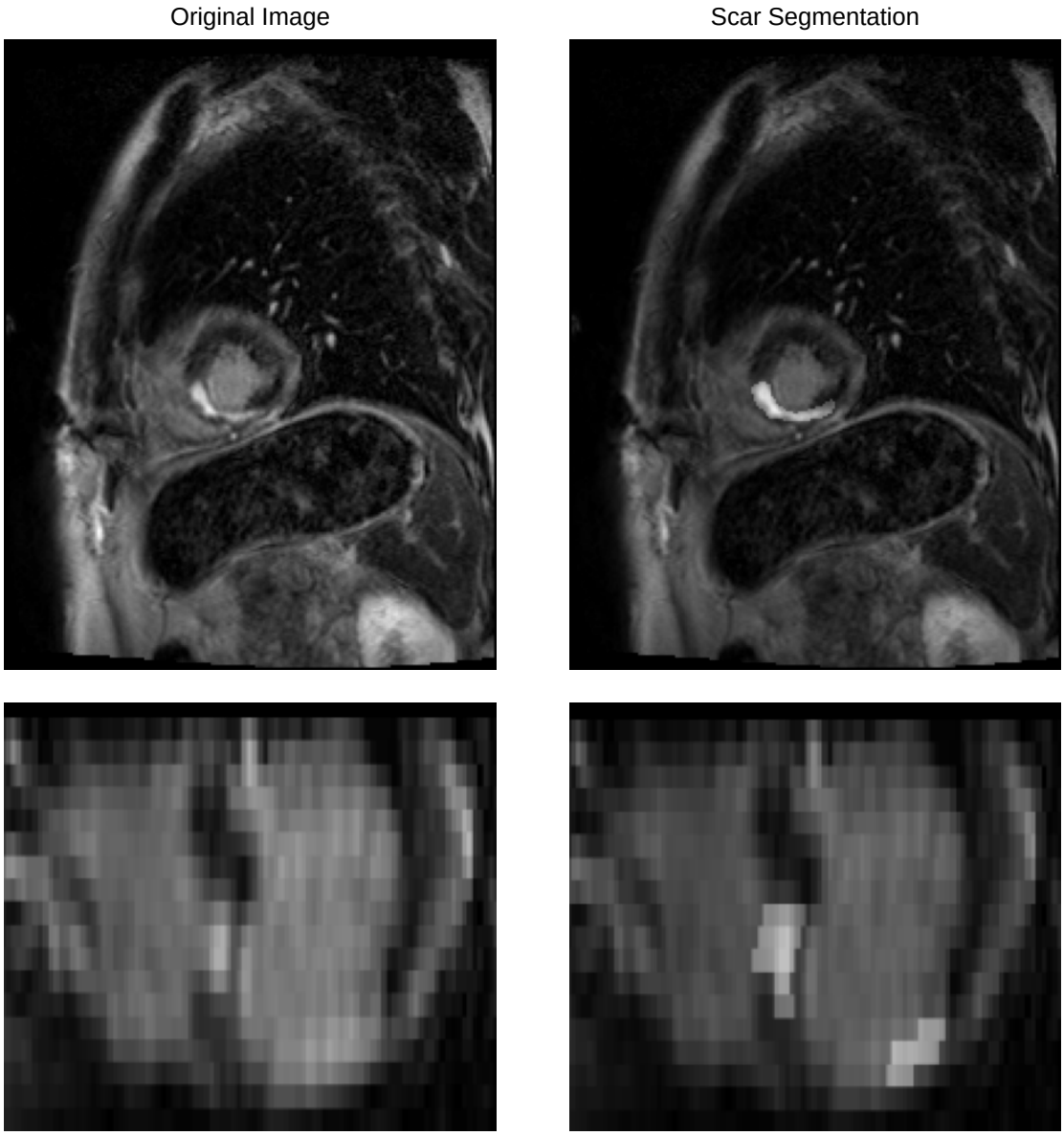


Fig. 1.8.: Example of an LGE CMR image and associated LV myocardial scar segmentation. The scar region can be identified as the hyperenhanced region within the myocardium. The current limitation includes the large slice resolution and lower contrast at the interface between subendocardial scar and blood. The top row: the short-axis view (z-axis); the second row: the y-axis of the short-axis view.

administration. This technique allows the assessment of the scar region comparable to delayed enhanced CMR [Mahnken, 2005]. With cardiac CT imaging, the myocardial scar assessment can also be done via the LV wall thinning (< 5 mm) [Komatsu, 2013], which is the result of the myocardial remodelling as detailed in section 1.2.2.

Figure 1.9 shows an exemplary cardiac CT image and the thinning regions of the LV. The majority of our work described in this thesis will focus on studying the relation between the LV wall thinning and scar-related VA, which is possible thanks to the availability of the large imaging and clinical database.

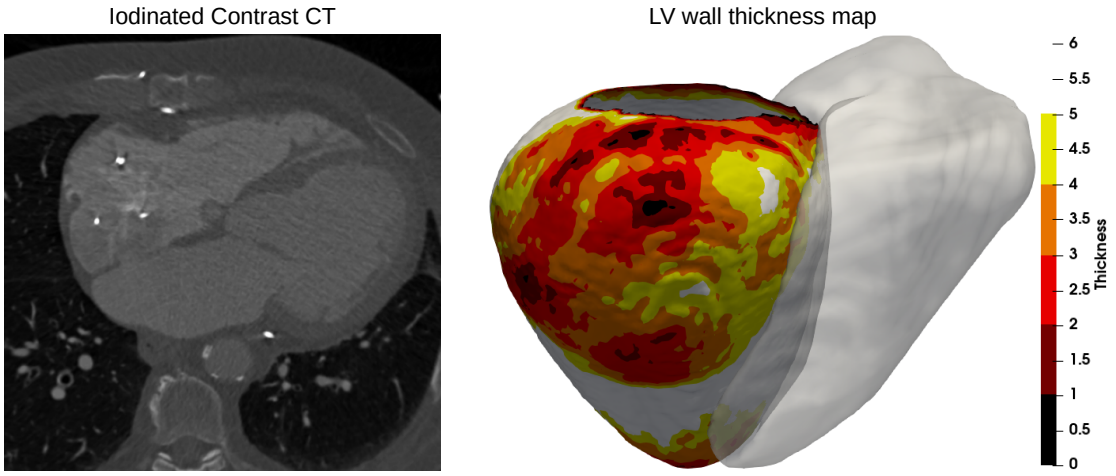


Fig. 1.9.: An example of the iodinated contrast cardiac CT image and the 3D LV wall thickness map. The heterogeneity and the extent of myocardial can be assessed through the thinning of the LV wall (thickness < 5mm).

1.3 Methodological Context

1.3.1 Deep Learning

Deep Learning (DL) methods have been studied and used very extensively in the medical field, from image processing to diagnostic and outcome prediction. In this work, we utilise a 'supervised' DL model in order to complete two main tasks: the segmentation of the ventricle in cardiac imaging data (see chapter 2 and 3); and, the classification task for predicting the presence/absence of VA from input features (in chapter 4 and 5).

1.3.1.1 Basic Concept

Deep Learning is a sub-branch of Artificial Intelligence (AI), figure 1.10.A. While within the family tree of AI, DL belongs to several layers of parent branches and categories, DL can be ultimately summarised as a model that follows the neural network structure. The neural network model simulates the structure of the biological neuron, where the information is passed through the network of the interconnecting neurons, as shown in figure 1.10.B.

The information is passed from the input through the 'nodes' (each containing different weights) of each layers, to produce a desired output at the end. This first step is called 'feed forward'. As a part of the AI family, the DL can learn from the mistakes of the output by 'optimising' the 'cost' or 'loss', calculated as the difference between the current output and the 'ground truth'. This process is called 'supervised' 'training' or 'fitting'. The optimisation changes the weights in each nodes by taking into account the gradients

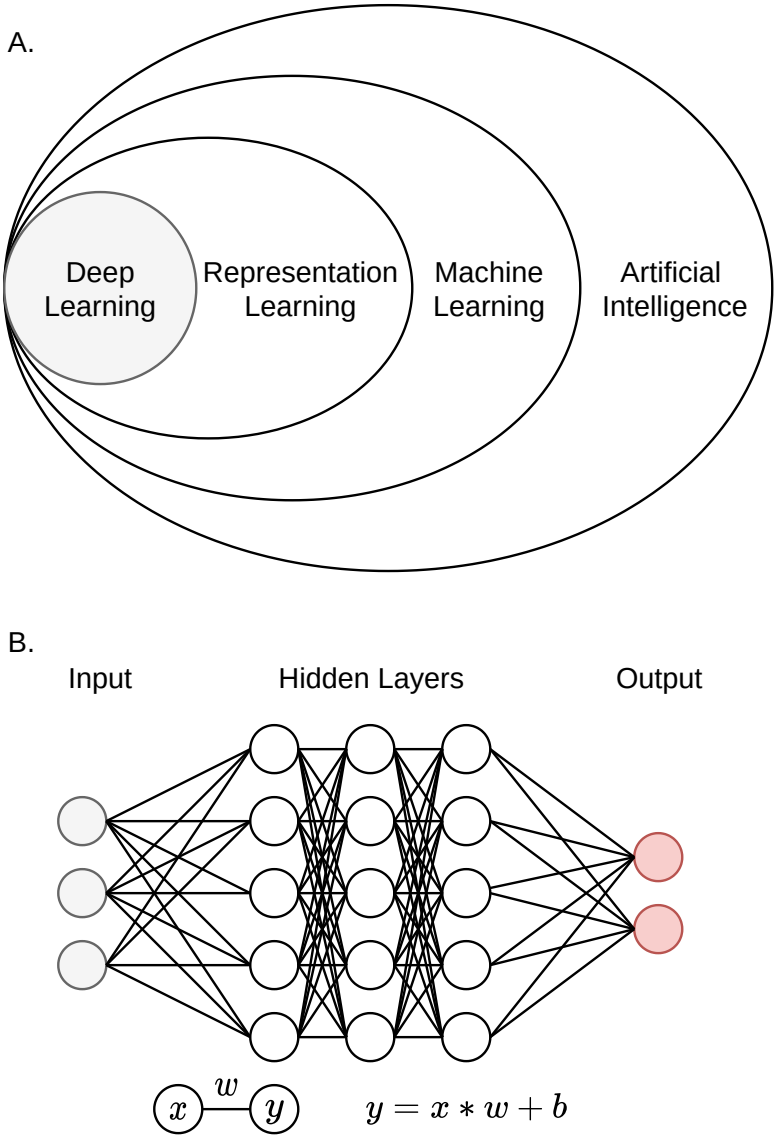


Fig. 1.10.: A. The Venn diagram of the deep learning family tree showing that DL belongs as a subcategory of the representation learning, machine learning and artificial intelligence (illustration adapted from [Goodfellow, 2016]). B. An example of a fully connected DL model. The fully connected layer produces an output y by multiplying the input x with the weight w before adding the bias b.

of the loss, which are backpropagated through each layers. This is often referred to as the 'backward' step of the DL model. The optimisation is done through the training input-output pairs to complete one 'epoch' (when all the training pairs were used).

As the number of the training epochs grows, the model becomes more and more adapted to the training data and can evolve to a point where the model cannot generalise to other data, which is known as 'overfitting'. To prevent overfitting, the model is 'validated' by feeding forward and calculating the cost of a separate set of data, not used for the optimisation. Most commonly, the validation is done after every training epoch, to gauge the performance of the model. The analysis of the validation loss can be used to control the 'learning rate', the amplitude of weights change during the optimisation, as well as to dynamically control when the training should end with 'early stopping'. Logically, if the validation score does not improve after a certain amount of training epochs then the lowering the learning rate might allow the access to the local minima, thus improving the generalisation of the model leading to better validation. On the other hand, if even after reducing the learning rate the validation score still does not improve, the training should be terminated, as further training would only lead to higher degree of overfitting.

Following the previous step, the concept of overfitting can, as well, be extended to the validation dataset. As the model is trained and stopped at an exact point that produces the best validation score, the model bias toward the validation data is unavoidable. To resolve this problem, another set of data is usually preserved until the end of the training for 'testing' or 'evaluation' of the 'generalisability' of the trained model.

In spite of this, the bias relating to the testing dataset can not be ruled out, since the curation of the truly generalised and unbiased testing dataset is not trivial and might be unfeasible in practice. Therefore, it is crucial to acknowledge the dataset bias when comparing and reporting the result of the model. In this work, an extensive testing and validation is done to study the viability of our DL model. In chapter 4, we used a bootstrapping cross-validation method, as well as a statistical analysis method, to ensure the fair comparison between the propose DL model and the clinical baselines. Similarly, in chapter 5, the cross-validation method was used.

1.3.1.2 Image Segmentation

In medical image processing, image segmentation is a processing step that draws the delineation of the targets, for example specific organs or lesions. Through the segmentation, the direct diagnosis, such as the presence/absence of lesion or cancer, can be done. Moreover, it also a crucial step in feature extraction of the organ or lesion, thus allowing us to perform further downstream analysis and predictions.

In the current clinical practice, the segmentation of medical images can be done in any of the following three approaches: manual; semi-automatic; and, fully automatic. The manual segmentation is done by the expert and the boundaries of the target have to be drawn by hand. This can be regarded as the gold standard method and is a necessary step to generate the ground truth, which permits the validation of newly developed semi-automatic and automatic methods. Nonetheless, this method is time-consuming, especially when considering the 3D or 3D+time image data. Moreover, the reproducibility of the manual segmentation is also subjected to the expert knowledge and experience, as well as the intra- and inter-expert variability. With this context, the intra-expert variability refers to the difference produced by a single expert, and the inter-expert variability refers to the difference between the segmentation result obtained by one expert compared to the result of a second expert.

The semi-automatic segmentation requires a minimal user interaction. Usually, it is in the form of landmarks or region of interest selection, from which an automatic algorithm is used to segment the target. In contrast, the fully automatic segmentation is capable of segmenting the entire dataset without any interaction from the user. While the two automatic methods have better reproducibility compared to the manual segmentation, assuming the models are robust, it is worth noting that the segmentation algorithm or the training dataset of the two methods might contain intrinsic biases related to the segmentation logic or the ground truth labels.

In recent developments, the DL approach, and more specifically the convolutional neural network (CNN), have taken over as the main model in image segmentation, thanks to its ability to process natural image data. Among the numerous architectures, the UNet architecture is, undeniably, one of the most revolutionising convolutional network architecture in medical image segmentation [Olaf, 2015]. The UNet model follows the autoencoder architecture, with an encoder and a decoder block, which are interconnected through a skip connection, as shown in Figure 1.11.

In our work dedicated to VA classification, the image segmentation is the key step to correctly quantify the LV scar region. In chapter 2, we devised the methods using the modified version of the UNet model to segment the LV, the RV and the LV scar regions in the LGE CMR images dataset. In chapter 3, a Dual UNet model (which was pre-trained in another project) was used as a key step to automatise the pipeline and generate the input dataset for the VA classification in the chapter 4 and 5.

1.3.1.3 Image Classification

As the name suggests, the image classification is a task that allocate the input image into a certain class (or label), according to the specific or class-defining features of the input.

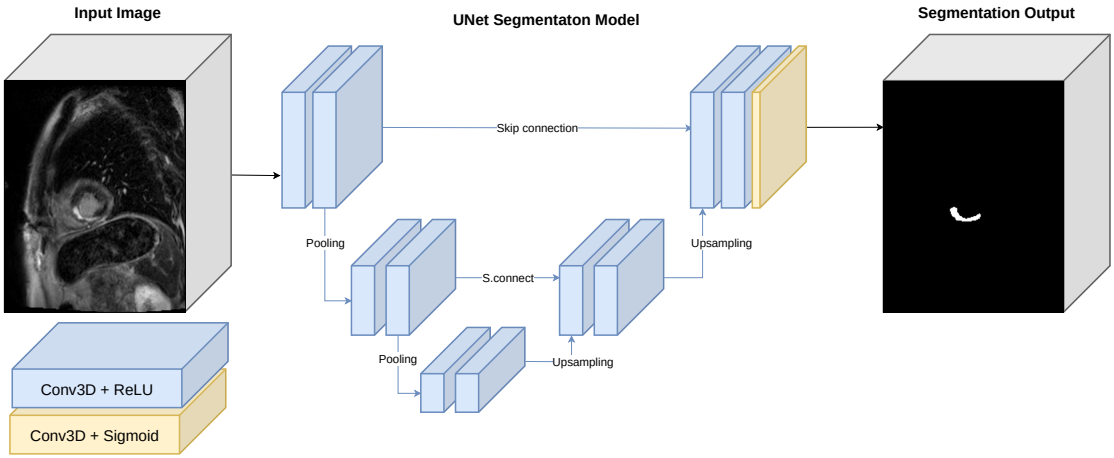


Fig. 1.11.: UNet model architecture for image segmentation. The image is fed through a series of U-shape convolutional blocks with decreasing (via the pooling layers) and increasing resolutions (via the upsampling layer). The skip connections allow the aggregation step and prevent the loss of the features from higher resolution.

In medical image processing, the classification can involve menial task such as image slice sorting, or more complex diagnostic applications such as lesion, cancer or other disease classification.

In clinical practice, these tasks would require different level of skill and experience to be correctly accomplished. As shown in Figure 1.12.A, the features have to be extracted from the medical image in order to be used for the classification either by the established statistical model (i.e. the cutoff value, etc) or by an ML classifier model. Using the DL approach, the technical setup for any classification task is the same, where only the input (image data) and output (the corresponding class) is needed. As in theory, the model optimisation of the model would automatically include: the pre-processing step; feature extraction; and, selection step from the input medical image. The feature extraction and selection through optimisation plays an important role in the exceptional performance of the DL model in image classification. Figure 1.12.B shows an abstract of the commonly used CNN model for the classification task.

This simple setup is also the pitfall of the DL approach, that is the 'blackbox'. In short, as the necessary steps allowing the image classification are automatised by the DL model, the model becomes un-human-readable, thus generating scepticism towards the model prediction. The issue is further intensified and not negligible in medical field. To tackle this problem, explainable learning is proposed as a sub-branch of DL, which aims to recover the features of the input that explain the model prediction.

To increase model transparency, a higher level feature extraction can be employed on the original medical image, such as the target segmentation. For the VA classification model, our attempt is to segment the scar region in the LGE images (see chapter 2) and

in cardiac CT images in chapter 3. However, limited by the number of available LGE CMR data and the segmentation accuracy of the segmentation result in scar region, we focus our attention in building the VA classification model from the LV wall thinning in the CT data, as described in chapter 4 and 5.

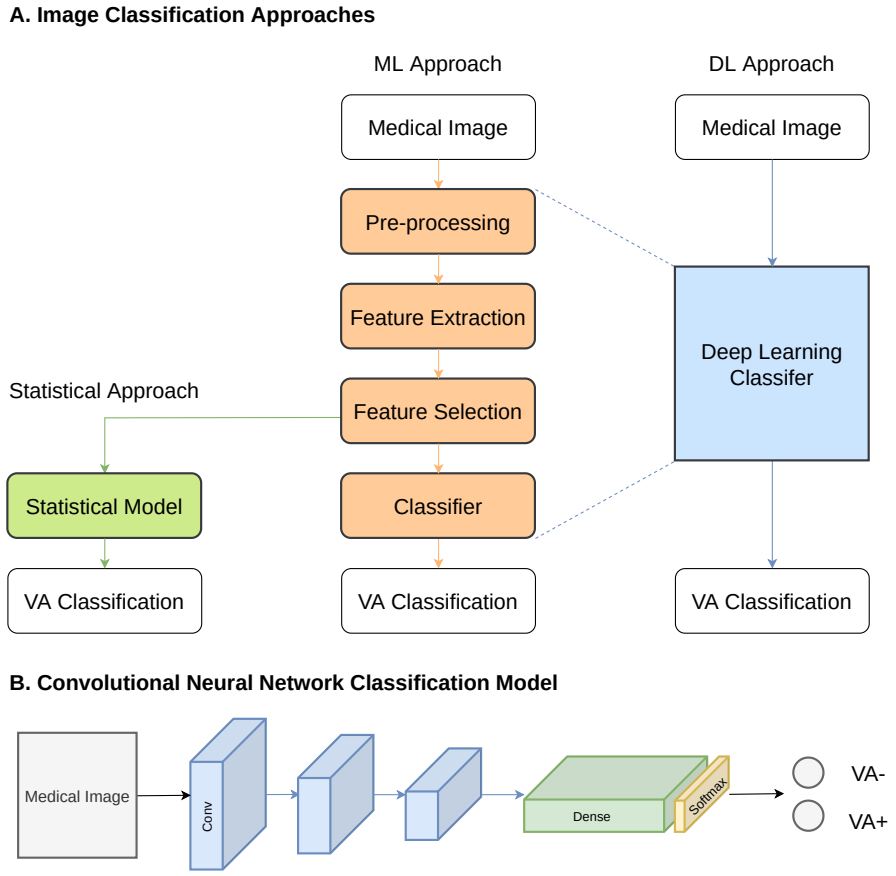


Fig. 1.12.: A. Diagram comparing different medical image classification approaches. (Illustration adapted from [Suzuki, 2017]). B. The CNN classifier network: an input medical image is fed through the succession of convolutional layers and followed by the dense/fully connected blocks and the softmax activation to produce the classification output.

1.3.1.4 Explainable Learning

In the recent years, explainable learning methods have seen a great resurgence and are used for building AI models. In particular, for the less transparent DL model, which usage increases and infiltrates into different fields. The goal of explainable learning is to give transparency to the inner work of the model and to provide human understandable explanation of the model decision. The explanation also allows the evaluation of the model fairness or the potential bias in the training of the model, which are crucial to increase understanding and trustworthiness of the AI model.

Under the umbrella of Explainable Learning, two terms are often used interchangeably 'explainable' and 'interpretable', although the distinction between the two terms is

still unclear even among the experts [Gilpin, 2018]. According to the author of a relevant study [Gilpin, 2018], the explainable AI is defined as the summary of all reasons for the behaviour or the insights of the causes of the model decisions, while the interpretable AI refers to the descriptions of the internals system of the AI model in a human understandable way. In short, the explainable AI explains why the model made its prediction but not how, and the interpretable AI explains how the model made its prediction but not why.

This semantic difference between the two methods, although fuzzy, serves as the foundation of the calculation method, where the difference is more distinctive. In this thesis, we used the explainable method, GradCAM++, detailed in chapter 4 and the interpretable methods, LIME (local interpretable model agnostic explanation) and integrated gradient, in chapter 5, respectively. While the calculations involved in the two methods is quite distinctive, the outputs of both methods are presented in a similar fashion, as shown in Figure 1.13. In medical image classification, as in the context of this work, the output is presented as the heatmap overlayed on the top of the original input. The heatmap explains the prediction by placing higher value on the "decisive" region. The calculation details and interpretation of the respective explanation methods are presented in chapter 4 and 5, respectively.

The results obtained in chapter 4, via the explainable method, allows us to solidify our theory that the thinning regions of the LV are the main factors permitting the DL model to accurately predict the presence of the VA. Nonetheless, the output would cover the entire thinning region, as shown in Figure 1.13, which does not provide much improvement to the therapy (i.e., the ablation), where specific characteristics of the arrhythmogenic substrate within the scar region are needed. In chapter 5, the application of the interpretable methods allows us to generate a heatmap that is more specific to the regions within/adjacent to the scar regions.

To some extent, the difference between the outputs, illustrates the distinction of the explainability and interpretability AI. As the prior highlights the entire scar regions to explain "why" the model made its prediction, the later gives specific scores to different part of the scar regions to explain "how" the model arrived at its prediction.

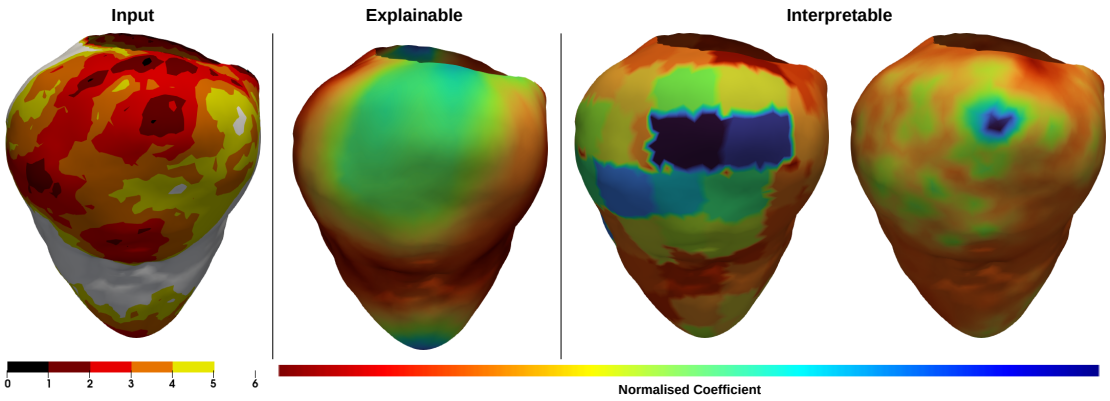


Fig. 1.13.: Example of explainability and interpretability outputs for a VA+ classification from the DL models, in chapter 4 and 5. The heatmap defines the explainable/interpretable regions, by placing higher coefficient on the specific regions.
As shown, the explainability method produced smoother heatmap compared to the interpretability method, where the contrast between local coefficients is more distinctive. While the explainablility map could assign different local coefficient values within the "decisive" region, the region is usually read as a whole. In layman terms, this could be read as “the thinning regions (the darker region) of the input is the reason “why” the model classified the input as VA+”.
In contrast, the interpretability method is based more heavily on the analysis of the local coefficient values to explain “how” the model reached its prediction; thus it can be read as “a combination of these local regions that allowed the model to reach its prediction”.

2

Automatic processing of cardiac post-infarct MR images

Contents

As mentioned in chapter 1, CMR has excellent contrast and capability to evaluate the location and extent of the myocardial scar. Thus, such scar images can be used as the base for our automatic VA prediction pipeline. However, the key step to a robust automatic image processing pipeline is the image segmentation step. As discussed in 1, one issue is that clinical CMR images are often limited by a relatively low resolution due to their relatively large slice thickness, as well as by a poor reconstruction consistency between MR scanners from various vendors, along with different local configurations and imaging protocols.

In this chapter, we specifically focused on developing an automatic pipeline to analyse the LV myocardial scar from clinical LGE MR images. Having a limited LGE imaging dataset available, we first explored in section 2.2 the possibility of integrating a multi-modality dataset, either from different imaging centres or by image augmentation method. However, although our method is proven efficient for the LV epicardium and endocardium segmentation, it performed poorly for the scar segmentation.

As a result, we then refocused on improving the scar segmentation accuracy using a mono-modality dataset. To do so, in the work described in section 2.3, we particularly introduced additional losses and training constrains to the scar region.

Figure 2.1 summarises the studies investigated in this chapter.

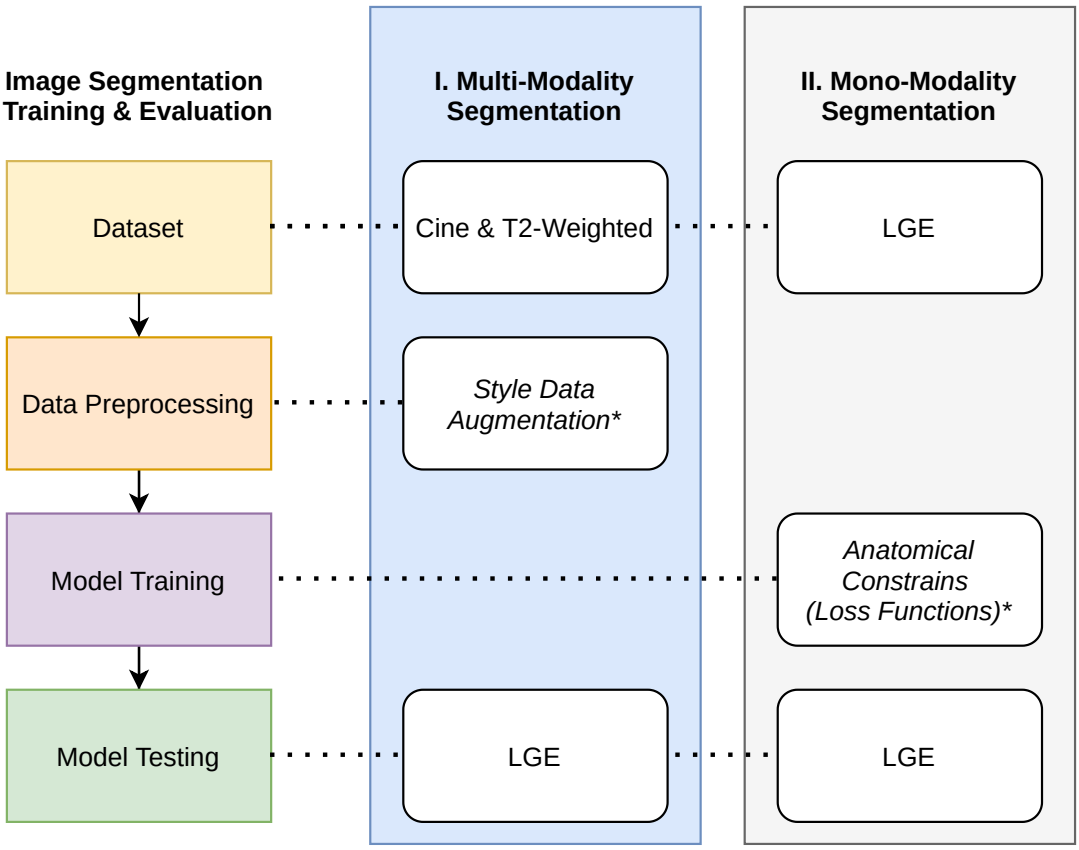


Fig. 2.1.: Overview of the studies investigated in this chapter. (*) indicates the main contribution of each section.

Our main contributions in this chapter are the following:

- We proposed the style data augmentation (SDA) method, which allows us to segment unseen modality images.
- We demonstrated that our DL segmentation model can be trained with multiple modality image.
- We demonstrated that the SDA method can improve structural segmentation accuracy even with mono-modality dataset.
- We proposed a TCL-Network model which produces fast and robust segmentation of CMR images.

- We proposed a new distance loss to apply additional constrain to the dice loss, further increasing the segmentation accuracy.

- We proposed a position loss to integrate anatomical constrains into the scar, as well as the epicardium and endocardium mask of the LV, respectively.

- We proposed a balanced-type of loss function with weights calculation to dynamically update the weights distribution for each loss functions during model optimisation.

- We proved that multi-channel segmentation model of the scar, epicardium and endocardium, allowing us to perform passive regulation of scar segmentation, which further increases its accuracy.

2.1 Context

2.1.1 Imaging Modality

The term modality refers to the specific quality or attribute shared between the item belong to the same modal. In the context of medical imaging, the images recorded using a certain modality would share the same attributes, such as: image resolution, quality of contrast, noise ratio, or the framing of the output image. For instance, figure 2.2 illustrates cardiac images obtained by three different modalities: echocardiography, CT and MRI. The same imaging technique, for instance MRI, can also produce different image modalities using different acquisition methods. In the context of this chapter, we use 3D CMR images of different qualities, depending on the contrast agent or the balanced weights employed during image reconstruction.

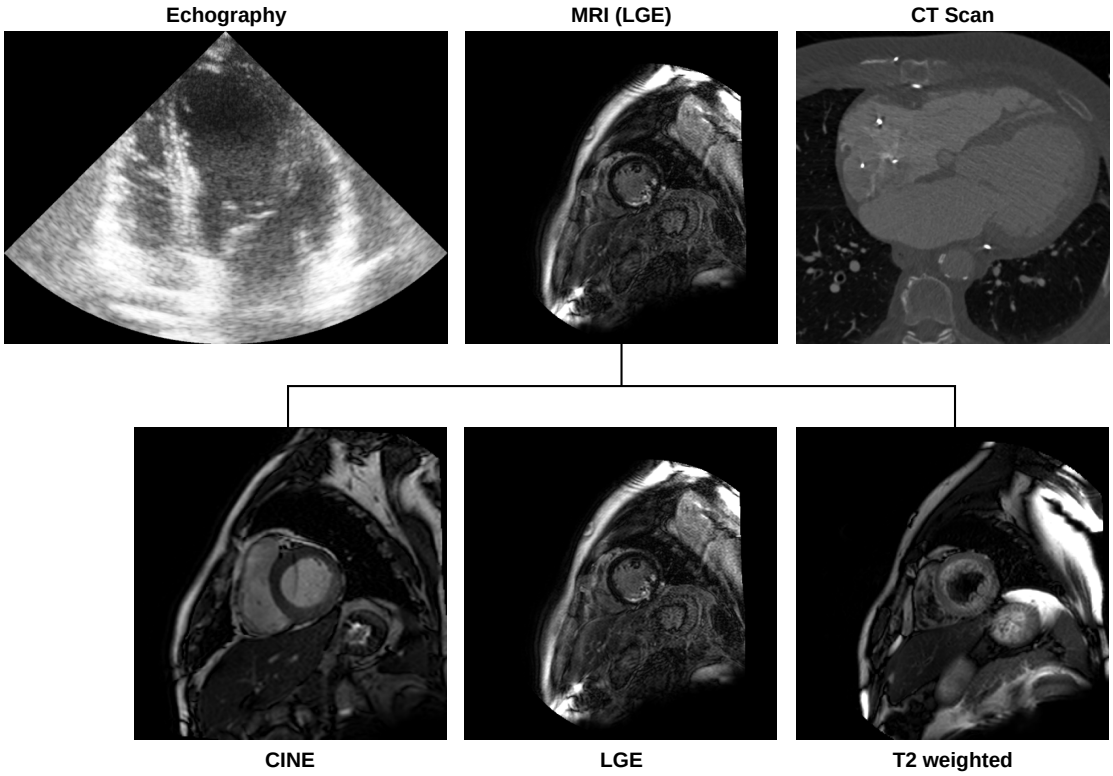


Fig. 2.2.: Examples of different cardiac imaging modalities. Top row (from left to right): Echography (Apical 4 chamber view), MRI (LGE, short axis view) and CT scan (trans axial view). Bottom row (from left to right): Cine MRI, LGE MRI and T2-weighted MRI.

2.2 Multi-Modality MRI Segmentation

This chapter is based on the following paper:
Style Data Augmentation for Robust Segmentation of Multi-Modality Cardiac MRI
Ly B., Cochet H., Sermesant M. STACOM 2019 - 10th Workshop on Statistical Atlases and Computational Modelling of the Heart, Oct 2019, Shenzhen, China.

2.2.1 Introduction

The combination of different MRI sequences, along with signal weighting techniques and contrast agents methods currently used in CMR, can produce a diverse modality and quality output image. Although each imaging technique yields exploitable results, the variation in image contrast can be detrimental for the development of automatic semantic segmentation tools in medical imaging.

To address the diversity of image quality in machine learning (ML) segmentation task, Seeböck et al. used an unpaired modality transfer generator network, CycleGAN [Zhu, 2017] that was able to reduce the variability in their multi-centre dataset study [Seeböck, 2019]. On the other hand, Isensee et al. proposed a nnU-Net (no-new U-Net) algorithm to automatically generate a CNN pipeline, optimised for each specific dataset [Isensee, 2019]. However, these methods require a sufficiently large mono-modality dataset and were specifically built to be used for mono-modality segmentation.

In this study, we used an alternative approach to this problem. We design a data augmentation method to train a single Deep Learning model to be robust to multi-modality input, including the modality that it has not seen during training. The style data augmentation introduces diversity of image contrast to the training dataset, with the goal to prevent the model from overfitting toward the training images' modality, and to divert the ML network's attention to the fundamental geometry features of the segmentation target. Here, we based such method on the theory that despite having a different contrast, the organ's geometry features are consistent within an MR imaging modality.

We integrated the 3D convolutional neural network for the segmentation, which incorporates the features in 3D space for the segmentation [Çiçek, 2016]. Nonetheless, the 3D image segmentation can be costly in terms of both memory usage and computation time, which can also increase exponentially as the network grows more complex. We have designed two strategies to combat this problem. Firstly, we proposed a minimalist U-Net inspired network, tailored to accelerate the convergence speed. Secondly, we adopt the dual network strategy [Jia, 2018], which allows us to segment high resolution targets.

2.2.2 Methods

2.2.2.1 Thresholded Connection Layer Network

We proposed the segmentation convolutional neural network called thresholded connection layer or TCL-Net. The network's architecture is shown in figure 2.3. The network is an iteration of the U-Net architecture, originally proposed by Ronneberger et al [Olaf, 2015]. As such, the network follows the same U-shape design and is made up of an encoder and a decoder portion.

The architecture of TCL-Net exploits the segmentation network's ultimate objectives, which are also meant to eliminate non-target pixels and to highlight the target pixels of the input image. TCL-Net uses, as building unit, two consecutive, padded, $3 \times 3 \times 3$ convolutional layers, each followed by the normalisation and non-linear activation layer. For normalisation, we use the instance normalisation, since the training is done with a single-input batch. We used the LeakyReLU [Maas, 2013] as the activation function of the encoder unit and of the first convolutional layer of the decoder unit and the ThresholdedReLU as the activation of the decoder's second convolutional layer. At the end of each unit, a $2 \times 2 \times 2$ max pooling is applied to the output of the encoder's unit. Correspondingly, a $2 \times 2 \times 2$ upsampling is applied to the output of the decoder's unit.

Using the ThresholdedReLU, the network could eliminate the non-target pixel at every level, starting from the deepest. Additionally, instead of the original concatenation, we used the multiplication layer to merge the output of encoder with the decoder, where the network can preserve the elimination progress from the previous thresholding. Moreover, the multiplication operation also allows for the amplification or reduction of the value of the current output, thus affecting the elimination likelihood of each pixel in the next thresholded layer. Toward the end of TCL-Net, a sigmoid activation layer was added in order to narrow down the output values to $[0 - 1]$.

2.2.2.2 Dual U-Net Strategy

In this study, we implemented the dual U-Net strategy [Jia, 2018], where two U-Net models were trained independently but can be used consecutively in the segmentation pipeline. The first U-Net is trained to segment the target from low-resolution inputs, that is because the dimension of the original image has to be reduced in order to reduce memory consumption. Then, it is used to segment the dataset, where we used this result to crop the full resolution target. The cropped dataset is fed to the second U-Net for a higher resolution, resulting in a refined segmentation.

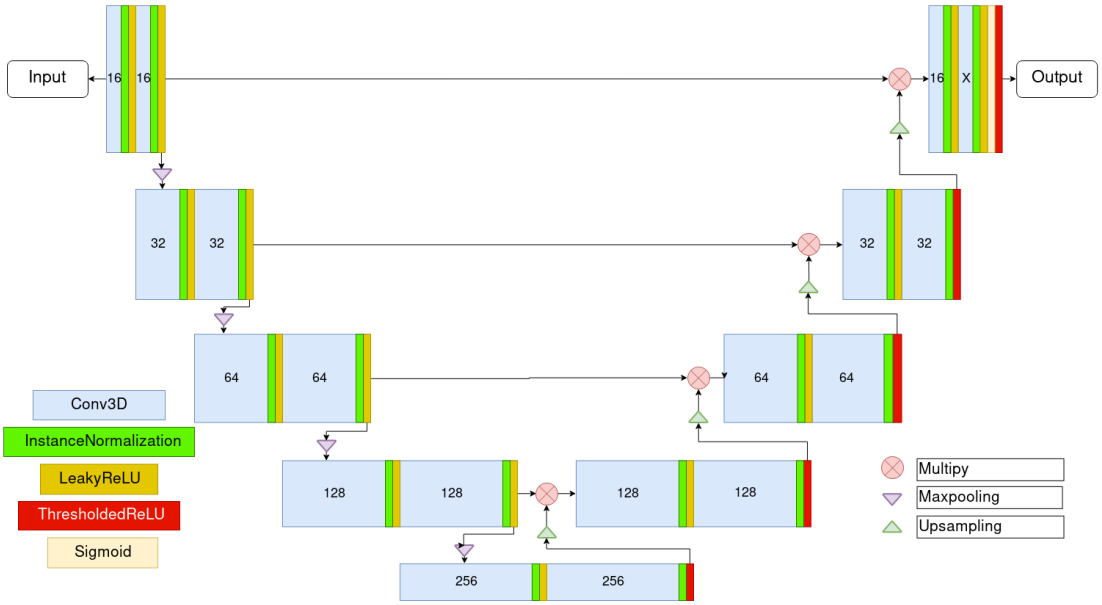


Fig. 2.3.: Thresholded Connection Layer network. *The number indicates the number of filters used at each convolutional layer. X denotes the number of channels of the input image.*

To crop the output of the first network, first, we rounded up the nonzero pixel to 1, then only the larger region of connected positive pixels is kept. To take into account the segmentation error, we applied the binary dilation transformation on the cluster using a $5 \times 5 \times 5$ spherical structure element. Finally, the original image was cropped using the bonding box of the dilated region.

In our study, we are interested in left ventricular segmentation from CMR images, specifically from LGE images, in which the myocardial scar is clearly visible. We used the first TCL-Net to locate the epicardium, and the second U-Net either to refine the segmentation of the target or to segment the scar (which was too small for the direct segmentation from the full-size image).

2.2.2.3 Style Data Augmentation

The style data augmentation strategy focuses on introducing contrast diversity in the training dataset. Our purpose is to prevent the model from overfitting to any single contrast of the training dataset and to divert the optimisation toward the fundamental geometry features of the target.

The image transformation algorithms were selected arbitrarily, as the goal is to simply increase the variety of the training images. For this study, we selected 5 transformation functions, as shown in figure 2.4, including: adaptive histogram equalisation [Hummel, 2008], Laplacien transformation, Sobel edge detection, invert intensity and histogram matching [Nyúl, 2000]. The histogram matching method was used to convert the original

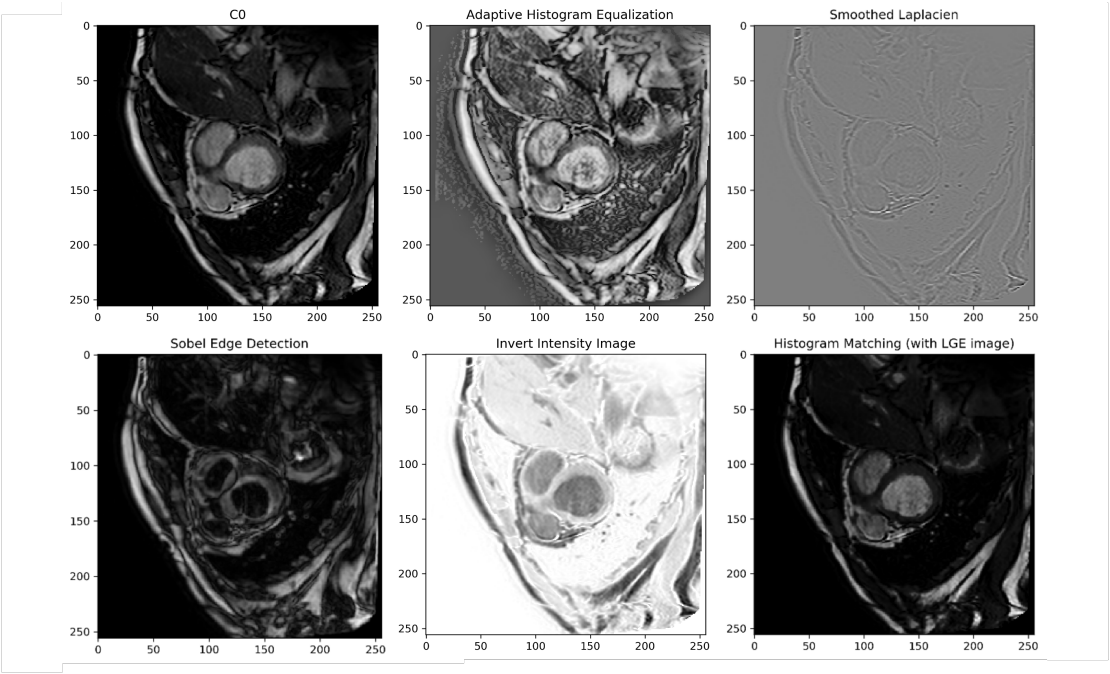


Fig. 2.4.: Different variations of input training images and the different image processing method. *C0 denotes the original steady-state pre-processing CMR image.*

training images (C0 and T2) toward the modality of the validation images (LGE). The complete detail of the datasets used for this study is described in subsection 2.2.3.1. These functions were applied to the normalised, eq 2.1, original image using the function provided by SimpleITK's python package [Lowekamp, 2013; Yaniv, 2018].

Our goal was to pre-train a segmentation model that is robust against any unseen-to-the-model image modality or quality. Our theory focuses on the geometry features; thus, it is only suited to be used where the target's shape is consistent.

2.2.2.4 Experimental setting

The network was trained using the batch size of one single 3D input array of $128 \times 128 \times 128$. The 3D images were interpolated to equalise the spacing of each dimension, thus the extracted data array could be closely corresponding to the physical size. We used the linear interpolation and nearest neighbour interpolation on the grayscale and mask image, respectively. The interpolated images was then re-sized to $128 \times 128 \times 128$. Finally, the array was normalised using linear normalisation function, eq 2.1, to bring the grayscale value to values between $[0 - 255]$.

$$X = X * \frac{255}{max(X) - min(X)} \tag{2.1}$$

To test the validity of this method, we did not apply any shape transformation for additional the augmentation or any complex pre-processing method on the validation or training image.

We used the initial learning rate at $1e-4$, which decayed by half each 5 epochs with no validation improvement. An early stop was also programmed after 20 epochs of no increase in validation performance. At each epoch, 100 images were chosen randomly from the training dataset to be used for training the network. To update the network, Adam optimisation was used along with the Dice loss, calculated using equation 2.2, where $\hat{Y}$ is the prediction mask, and Y is the manual label mask. This equation can preserve the gradient chain needed for the optimisation of the model. During the training, we also measured the Dice coefficient [Dice, 1945] between the prediction and the manual mask by applying the "half to even" round function on the segmentation output, as per the following equation 2.3. The round function would break the gradient chain, thus it cannot be used to update the model.

$$Dice_{Loss} = 1 - 2 * \frac{\sum(\hat{Y} * Y)}{\sum \hat{Y} + \sum Y} \tag{2.2}$$

$$Dice_{Coeff} = 2 * \frac{\sum(round(\hat{Y}) * Y)}{\sum \hat{Y} + \sum Y} \tag{2.3}$$

2.2.3 Evaluation on Clinical Data

2.2.3.1 Materials

For the multi-modality dataset, we used the dataset provided by the MS-CMRSeg 2019 segmentation challenge. The challenge's dataset consists of 135 3D images from 45 patients acquired using three modalities (i..e MR methods): T2-Weighted (T2), balanced-Steady State Free Procession (C0), and contrast-enhanced LGE, respectively, fig 2.5. The manual labelled masks were provided for the first 35 T2 images and C0 images, as well as for the first 5 LGE images. By applying our data augmentation method, we trained the network with 490 images with different variations of the original C0 and T2 images, and then validated the network with the 5 labelled LGE images.

To validate the effect of SDA on the homogeneous dataset, we used our local dataset, which consists of 119 mono-centre LGE-CMR images provided by IHU LYRIC. Since the dataset is mono-modality, we remove histogram matching from the SDA algorithms. The original dataset was first split in 9:1 ratio for training and validation, before the data

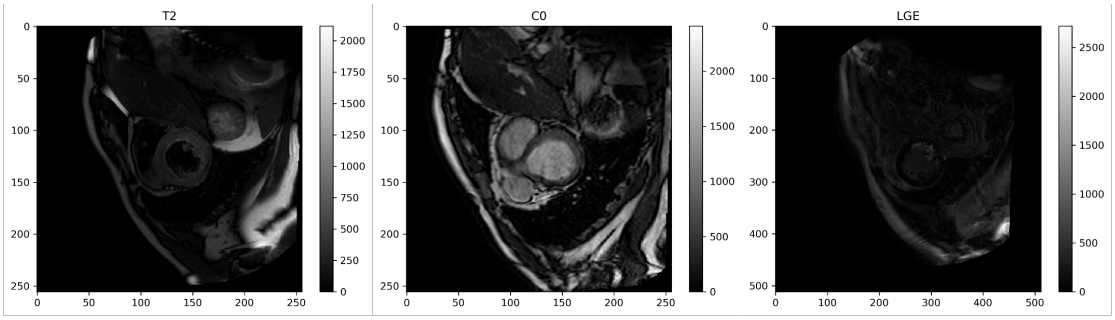


Fig. 2.5.: Short-axis view of T2, C0 and LGE MRI modality.

augmentation method was applied to the training images. We compared the mean of best validation scores from 5 different sets of validation images of the model trained with and without the SDA method.

2.2.3.2 Results

As shown in figure 2.6a, the SDA method increases considerably the segmentation performance of the epicardium in the first TCL-Net. The effectiveness of diversity in training dataset can also be observed in the result of the training using the combination of C0 and T2 images.

Figures 2.6b and 2.6c show the comparison between training and validation loss for each training dataset. We can observe that the model would quickly overfit with the mono modality datasets. The difference between the training and validation loss gets smaller with more diversity in the training dataset. The validation curve also appears more stable when training with both modalities input and the most stable when training with SDA method. The training with TCL Net shortens considerably the convergence time, as we can observe a stable validation and training loss in around 40th epoch for the training with SDA dataset, fig 2.6c.

When training with second TCL-Net, the validation Dice coefficient increases from 0.928 to 0.962. Figure 2.7 shows the validation segmentation output of the trained model using the SDA method with a single and dual TCL-Net, respectively. The segmentation seems to perform well and produces accurate segmentation in the region where there is no myocardial scar. Nevertheless; however, the model encounters difficulties at the scar regions, as pointed by the arrows in figure 2.7b and 2.7b.

When testing on the mono-centre and mono-modality dataset the SDA method did show improvement in validation Dice scores for the epicardium segmentation in the first TCL-Net, table 2.1. However, the method had an adverse effect on the myocardial scar segmentation in the second TCL-Net.

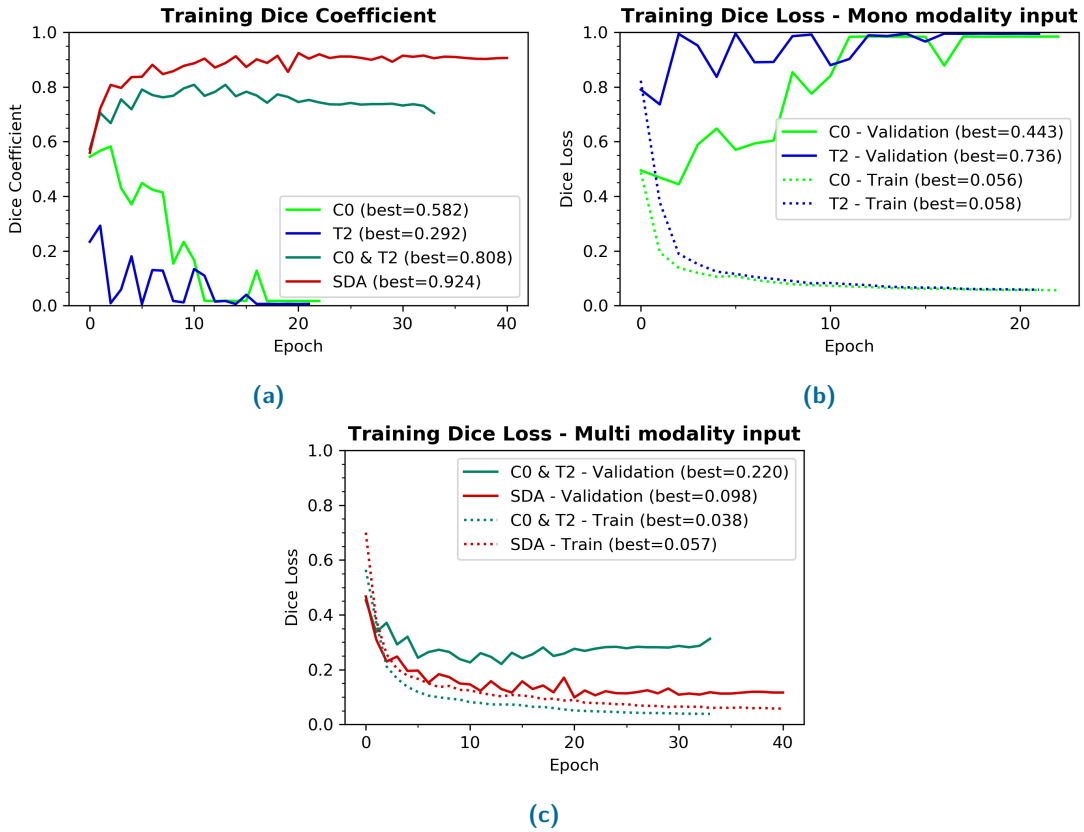
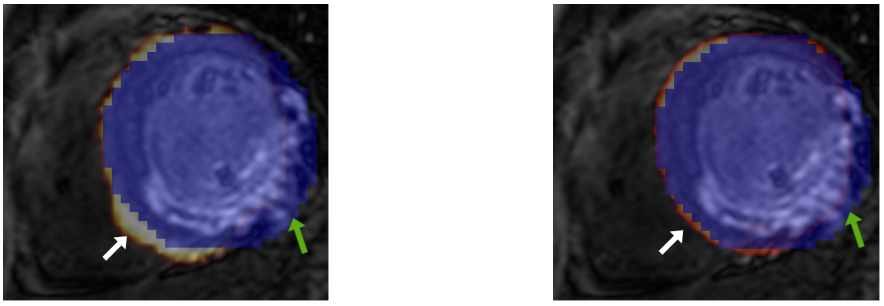


Fig. 2.6.: Validation result of epicardium segmentation of the first TCL-Net.

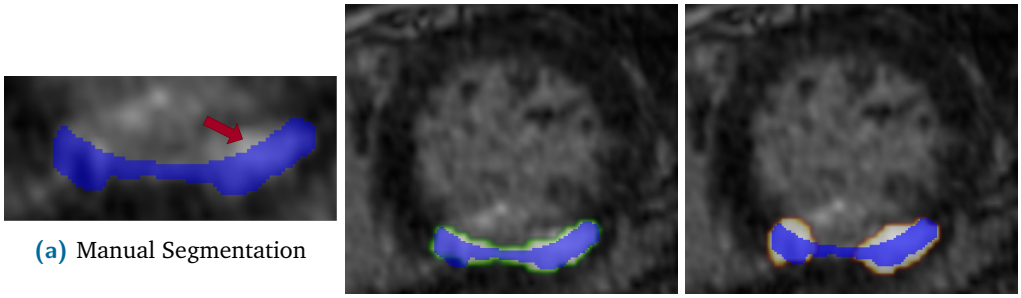
Tab. 2.1.: Validation Dice Loss 2.2 and Dice Coefficient 2.3 on our local data.

	First TCL-Net *(Epicardium)*		Second TCL-Net *(Scar)*	
	Dice Loss	Dice Coeff	Dice Loss	Dice Coeff
With SDA	0.1194 ± 0.0083	0.9055 ± 0.0114	0.5931 ± 0.0563	0.4444 ± 0.0615
Without SDA	0.1437 ± 0.0250	0.8741 ± 0.0280	0.5678 ± 0.06418	0.4626 ± 0.0707



(a) Segmentation with single TCL-Net **(b)** Segmentation with dual TCL-Net

Fig. 2.7.: Example of LGE epicardium segmentation using the model trained using the SDA method on C0 and T2 images. *Blue: Ground Truth; Orange: Segmentation Target.*



(a) Manual Segmentation **(b)** Without SDA **(c)** With SDA

Fig. 2.8.: Myocardial Scar Segmentation using dual U-Net strategy. *Blue: Manual Segmentation; Green: without SDA; Red: with SDA.*

Figure 2.8 shows the validation scar segmentation of the model trained with and without the SDA method. Despite the poor Dice coefficients, both models can adequately detect the scar region within the myocardium. Noted that as the targets become smaller, minor segmentation errors can become more pronounced in the Dice score. We can also note that the model trained with SDA dataset tends to over-estimate the predicted scar area, more so than the model trained with the original dataset, while also discounting the connection between the two scar regions.

2.2.4 Discussion

The image processing functions implemented in the SDA method create images with different contrast, defined border and geometric features, therefore making the method optimised for regular structures such as the epicardium. However, the method might not be suitable for the targets with heterogeneous structure, such as the myocardial scars.

Because the model can no longer depend on image contrast for the segmentation, it has to rely on the patterns of the target, such as the traces of the myocardium wall and the homogeneity of signal intensity of each structure. For instance, when training on C0 and

T2 dataset of MS-Seg challenge, the model does not work well on the scar region since it is only familiar with homogeneous myocardium zones in C0 and T2 images.

On the same note, the manual segmentation of the myocardial scar (which was used to calculation the loss) does not follow the same geometric or intensity homogeneity rules as the epicardium. As shown in figure 2.8a, the scar region does not include the entire area of the same intensity, since the upper area belongs to the cavity of the ventricle. The model trained with only the original image might be more familiar with this rule than the model trained with SDA, since it can depend more on specific contrast of the LGE modality during optimisation.

2.2.5 Conclusion

We proposed a data augmentation strategy that increases the accuracy of the segmentation and is invariant to the modality of the validation image. The SDA strategy forces the network to be independent from the input image modality and prevents overfitting to any specific contrast. By introducing a diverse style of input to the training data, SDA challenges the traditional concept of complex normalisation or equalisation of the dataset, in medical image segmentation. It also pushes the boundary of how powerful and flexible the convolutional neural network can be for the segmentation task. Using the proposed methods, the trained model can be used to segment input images with very little pre-processing work or intensity correction, while still producing good results.

2.3 Mono-Modality MRI - Scar Segmentation

As shown in the previous section, the scar segmentation is sensitive to the specific contrast of LGE images, thus preventing us to use the multi-modality image augmentation method proposed in the previous section. Therefore, in this section, we explored the possible method to improve the scar segmentation only from the mono-modality dataset, by focusing on the scar anatomical constrained and its relation to the epicardium and endocardium, respectively.

2.3.1 Methods

In the previous section, the segmentation models were only trained with the dice loss. In this section, we explore different losses and constrains that could be used to increase the robustness of the myocardial scar segmentation. To this end, we included the distance and position loss to the dice loss to train the model. Here we propose an automatic weights-updating method to dynamically adjust the weights distribution of the three losses. Finally, we also tested the multi-channel segmentation model as the passive anatomical constrain to improve the segmentation accuracy of the scar region.

2.3.1.1 Distance Loss

The dice algorithm calculates the loss based on the intersection between the two structures. This leads to the density bias as the optimisation is concentrated on the centre, which mostly lies in the target voxel and less on the border of the mask. To enforce the model optimisation on the border of the target, we applied the distance loss, inspired by the contour loss proposed by [Jia, 2018].

For the contour loss, these authors based the loss calculation as the normalised intersection between the contour of the segmented mask and the distance map. The distance map is calculated as the euclidean distance of each voxel from the ground truth binary mask. Therefore, the false segmentation further from the target will incur a higher cost. The original method proposed a contour clipping, either via value thresholding or by using a 3D Sobel filter.

For our study, to make the implementation simpler, we skipped this step and directly calculated the distance loss directly with the model output mask, which also penalised all false voxels. Moreover, we normalised the distance map to the [0-1] range, as shown in figure 2.9. From our experiment the normalised distance map limits the scale unbalance between the distance and the dice loss, thus making the model fitting more stable at the earlier epochs. The distance loss was calculated using the equation 2.4, where N is

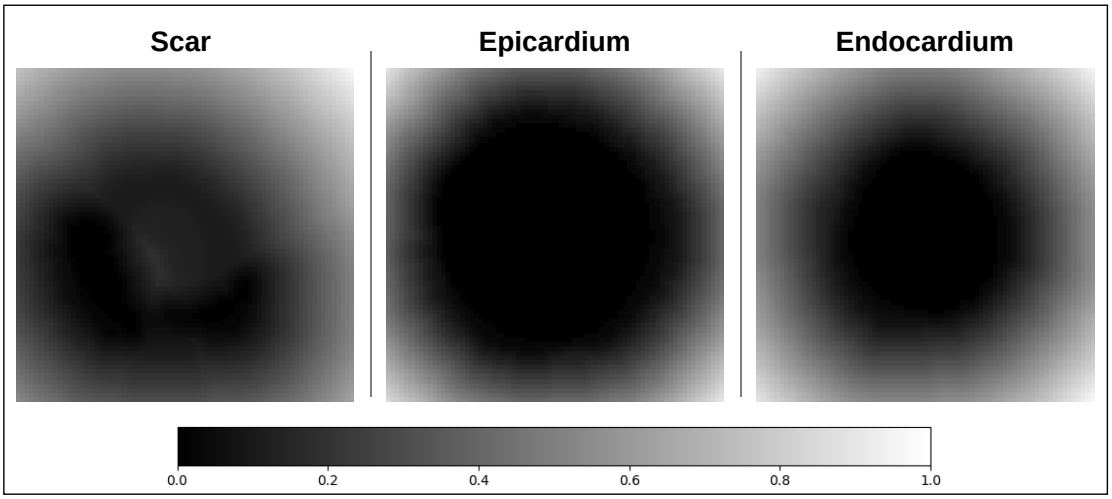


Fig. 2.9.: Normalised euclidean distance map of the LV scar, epicardium and endocardium.

the total number of voxel, $D(y)$ is the normalised euclidean distance map of the ground truth mask, and $\hat{y}$ is the segmentation mask.

$$L_{dist} = \frac{1}{N}\sum_{i}^{N}(D(y)_i * \hat{y}_i) \tag{2.4}$$

2.3.1.2 Position Loss

To further take into account the interaction between the different region of the LV, we introduced the position loss into the total loss function. Individual losses for each masks were proposed based on the anatomical constrained between the scar, epicardium and endocardium.

For instance, we know that the scar region must be contained within the myocardium. Thus, any scar zone predicted beyond the epicardial border or within the endocardium is penalised. The position loss for the scar was calculated using the equation 2.5, where $\hat{y}$ is the model segmented mask, $\overline{y}$ is the invert of the ground truth mask and y is the ground truth mask.

$$L_{scar} = \sum \hat{y} \cap \overline{y}_{myo} = \sum \hat{y} \cap (\overline{y}_{epi} + y_{endo}) \tag{2.5}$$

We calculated the position loss of the epicardium by enforcing it to cover the entirety of both the scar and endocardial region, using the equation 2.6. Inversely, we calculated the

position loss of the endocardium with the equation 2.7 by penalising any false positive segmented voxels over the scar region or outside of the epicardium.

$$L_{epi} = \sum(y_{endo} + y_{scar}) - (\hat{y}_{epi} \cap (y_{endo} + y_{scar}))) \tag{2.6}$$

$$L_{endo} = \sum \hat{y}_{endo} \cap (\overline{y}_{epi} + y_{scar}) \tag{2.7}$$

For this study, the position loss penalty masks were computed offline during data processing, although the masks can be easily and rapidly calculated during training. Similar to the calculation of the dice loss, we replaced the mask intersection operation using multiplication. The example of the penalty masks for the position loss of the scar, epicardium and endocardium are shown in figure 2.10.

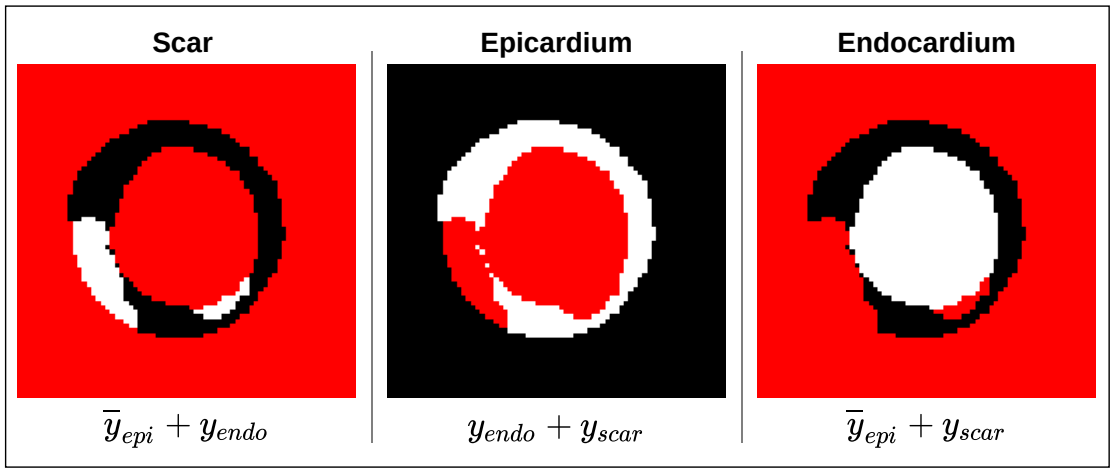


Fig. 2.10.: Position loss penalty map (red) and the corresponding mask (white). Note that while the scar and endocardium segmentation should not overlap the red area, the epicardium should cover the entire red area.

2.3.1.3 Dynamic Balance Weighted Loss

The total loss function used to train the segmentation model is shown in equation 2.8, where α, β and γ are the weights for the respective losses.

$$L_{total} = \alpha * L_{dice} + \beta * L_{dist} + \gamma * L_{pos} \tag{2.8}$$

The loss function weight allows us to adjust the order of magnitude of different losses. Moreover, it can also be used to designate the order of importance for each loss at the optimisation step. In this study, we proposed an automatic method to dynamically adjust (while the training progresses) the weight of each loss function with respect to the desired ratio. Figure 2.11, show the flowchart of the dynamic weights calculation.

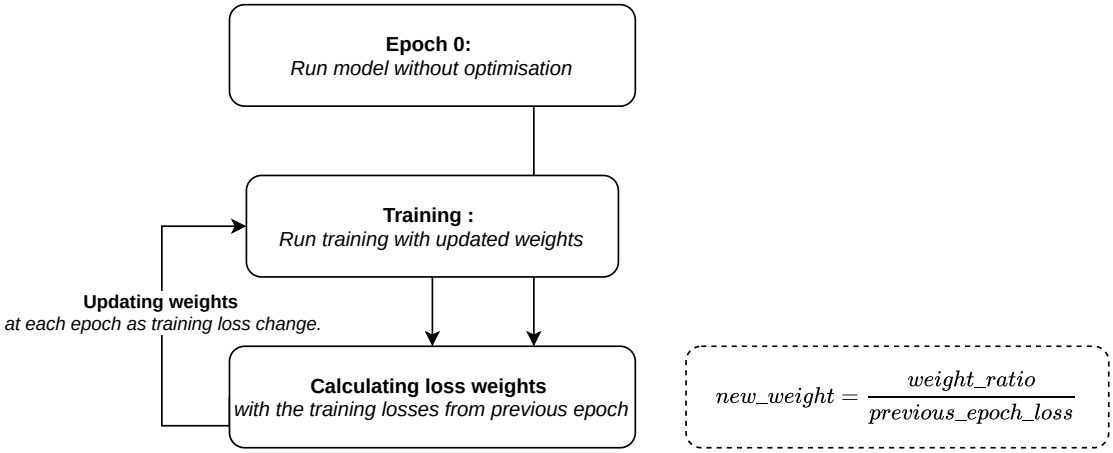


Fig. 2.11.: Dynamic balance loss function weights calculation.

For this project, we aimed to set weights ratio of the L_{total} in eq. 2.8 to $0.9 : 0.05 : 0.05$ for the L_{dice}, L_{dist} and L_{pos} , respectively. During the first epoch, we optimised the model by setting α β and γ using their respective ratios. The optimisation in the first epoch can be skipped since the loss functions weights are not adjusted. Through empirical testing in our study, the optimisation in the first epoch was tolerated and gave similar result. For each epoch, the new weights for each loss function were calculated by dividing the ratio with the mean loss from the previous epoch.

2.3.1.4 Multi-Channel Segmentation model

Finally, here we also proposed to study the passive constrain that can be imposed by using multi-channel segmentation model. Our hypothesis was that as the model has to learn to simultaneously segment relating structures, unenforced (i.e., without specific loss function) regularisation would be done to preserve the coherence between the outputs from each channel. For instance, the voxel should not be "activated" at scar and endocardium at the same time, despite sharing very similar input intensity. Moreover, the voxel must be activated for the epicardium channel, if it is activated for either scar or endocardium channel.

Multi-channel model is also faster to implement and reduces the resource needed in the pipeline, as the switch from mono-channel to multi-channel model can be as simple as changing the number of output of the final convolutional layer. Although the optimisation time may increase, because of the increase of the loss function calculation time. Figure 2.12 shows the multi-channel segmentation pipeline and loss calculations for this study.

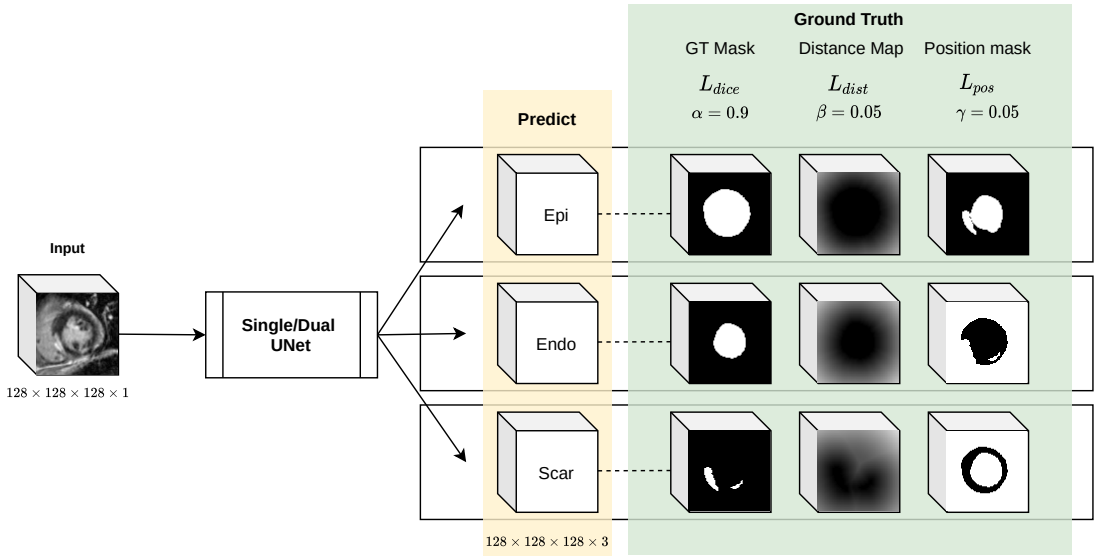


Fig. 2.12.: Multi-channel segmentation model with dice, distance and contour loss.

2.3.2 Evaluation on Mono-Modality

2.3.2.1 Experimental Setup

We test the effect of the new loss functions using random sub-sampling cross-validation. To maximise the training dataset, the dataset of 131 cases is randomly divided into 5 sets of training and testing data at a 9:1 ratio. The cross-validation is run on single or dual UNet model with different combinations of input masks and loss functions.

We used the TCL-Net model architecture as described in section 2.2, and the trained single-UNet model as the first model for the daul-UNet of the same category. For example, the single-UNet scar model trained with dice, distance and position loss was used to crop refine input to train the dual-UNet scar model of the same loss configuration. The models were trained with Adam optimiser, with the initial learning rate of $1e-4$, which was programmed to decrease by half every 10 epochs if no improvement of the validation dice score. The early stopping criteria were 20 epochs with no validation dice score improvement.

2.3.2.2 Results

The cross-validation results are shown in the table below 2.2. Overall, the dual multi-channel model trained with dice, distance and position loss achieved the best scar dice score at 0.586 ± 0.014. In contrast, the single model segmentation of the single scar channel with no anatomical constrains only achieved a 0.270 ± 0.176 dice coefficient. With the addition of the refined model, the single scar channel with only dice loss achieved

0.455 ± 0.037, which is still considerably lower than the best model. As the accuracy of the single scar model was not adequate, we substitute the first model of the dual scar model with the pre-trained epicardium model.

In general, the addition of both distance and position loss increase the scar segmentation accuracy, compared to the training with only the dice loss. Similarly, the multi-channel models also achieved better robustness compared to the single channel model trained using the same loss configuration.

Tab. 2.2.: Cross-validation results with 9:1 ratio of 131 LGE images. 5 fittings are done using random sub-sampling for each configuration. The rows are sorted by the scar dice score in descending order. (o). activated option; (-): deactivated option.

		Mask			Loss Function			
No.	**Dual-UNet**	**Scar**	**Epi**	**Endo**	L_{dice}	L_{dist}	L_{pos}	**Scar Dice Coef.**
1	o	o	o	o	o	o	o	0.586 ± 0.014
2	o	o	o	o	o	o	-	0.565 ± 0.023
3	-	o	o	o	o	o	o	0.550 ± 0.029
4	o	o	-	-	o	o	-	0.545 ± 0.023
5	-	o	o	o	o	o	-	0.535 ± 0.020
6	o	o	-	-	o	o	o	0.530 ± 0.035
7	-	o	-	-	o	o	-	0.495 ± 0.026
8	-	o	o	o	o	-	-	0.485 ± 0.040
9	-	o	-	-	o	o	o	0.477 ± 0.026
10	o	o	-	-	o	-	-	0.455 ± 0.037
11	o	o	o	o	o	-	-	0.415 ± 0.084
12	-	o	-	-	o	-	-	0.270 ± 0.176

2.3.3 Discussion

Although, the proposed method increased the dice score of the scar segmentation, the final segmentation accuracy is still considerably low, which translated into a poor output quality, as shown in figures 2.13 and 2.14. In figure 2.13, we can observe the failure of the automatic model to distinguish between the scar region and the endocardium. This can be partly attributed to the significant intensity inconsistency between the scar region of the top and the bottom row. Moreover, we can also note the lack of clear delineation between the endocardium, the myocardium, and the epicardium, respectively. The nature of the heterogeneous scar regions in LGE images can be contrary to the fundamental basis of the DL model, since the model is trained to recognise the repeated and consistent pattern from the inputs.

As shown in figure 2.14, the DL model can somehow detect and highlight the centre of the scar region, which likely contributed to the average dice coefficient 0.586. Nonetheless,

the delineation of segmented scar is still far from accurate, which either leave out the majority of ground truth scar undetected or overestimate it from voxels that actually belong to the healthy myocardium. In the context of this study, we kept the post-processing of the scar segmentation to a minimum, in order to better focus on investigating the DL model capacity. For this reason, there are small "noisy" voxels scattered across the image. The noise outside of epicardium can be easily removed. However, the noisy voxels within the epicardium cannot be simply removed, since these can very well belong to the scar region. This fact further contributes to the complexity of post-processing step in the scar segmentation.

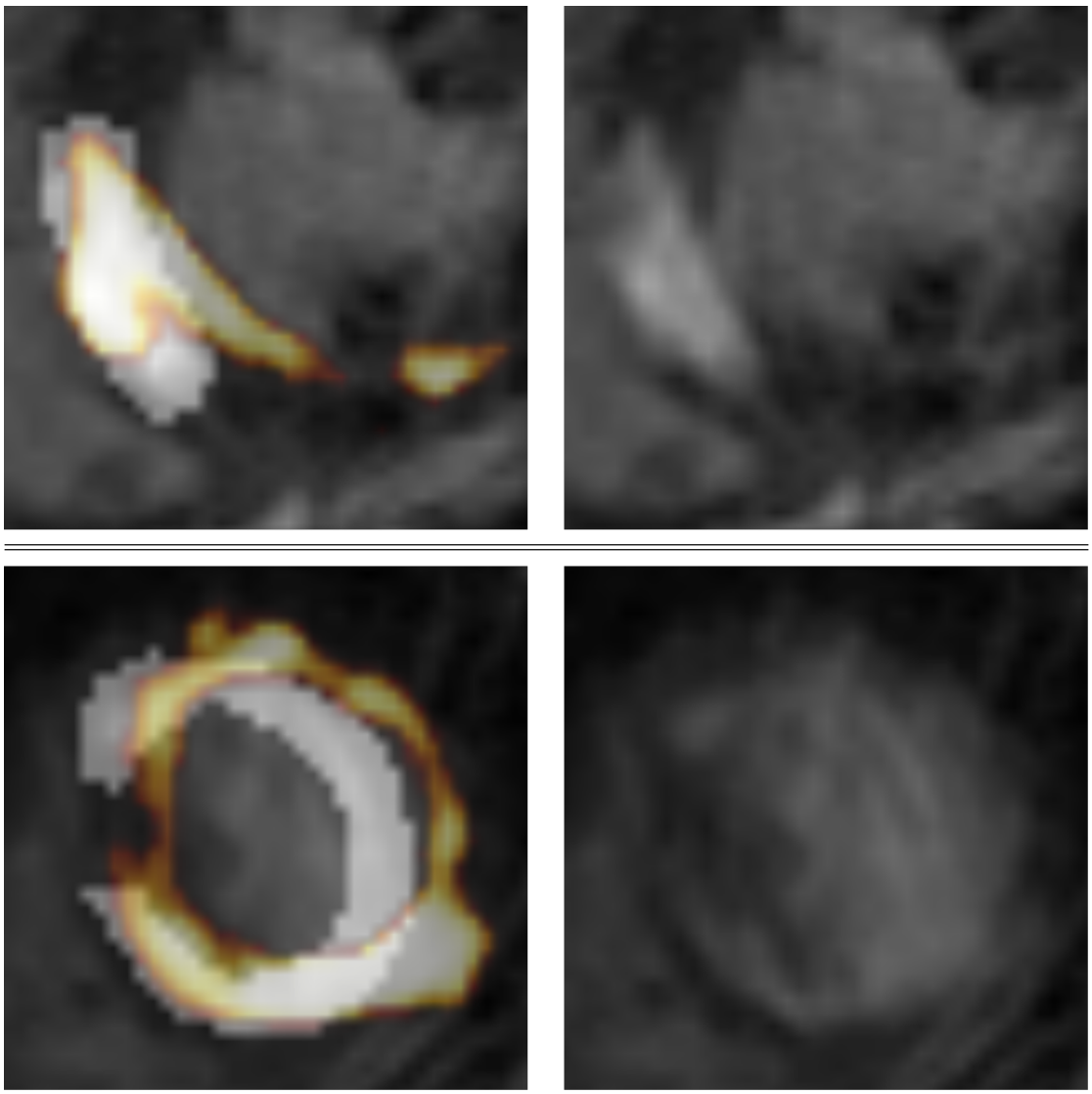


Fig. 2.13.: Comparison between automatically segmented scar and ground truth scar in a 2D slice. Orange: Automatic segmentation; White: Ground Truth.

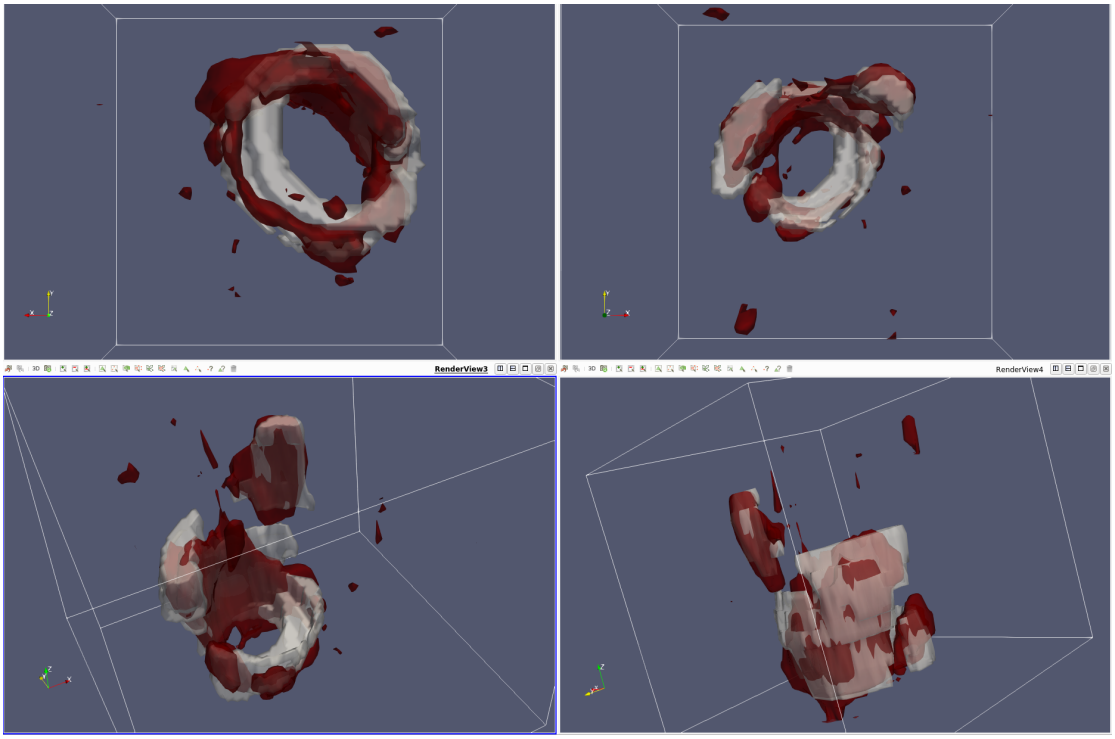


Fig. 2.14.: Comparison between automatic segmented and the ground truth myocardial scar in 3D mesh. Red: Automatic segmentation; White: Ground Truth.

2.3.4 Conclusion

In conclusion, we proved that the addition of the distance loss and position loss as the anatomical constrains along with the dice loss improve the accuracy of the scar segmentation. Cross-validation evidence also showed that multi-channel segmentation of scar, epicardium and endocardium allowed us to perform a passive regulation between their corresponding three masks, which increased the segmentation accuracy of the scar region.

Although the addition of the loss functions and the passive regulation from multi-class segmentation increases the scar segmentation accuracy, the scar segmentation is not sufficiently adequate for the quantitative or qualitative analysis. This could be attributed to the limited training dataset, as well as the low image quality and contrast of these clinical LGE images, which contributed to an unclear distinction of the scar and the endocardium. While further improvement can be made to better segment the scar region, the available dataset (comprised of 131 LGE images) would still prevent the appropriate training and evaluation of the eventual VA prediction pipeline. Therefore, to move forward with our main goal of developing a fully automatic method for the VA prediction from non-invasive imaging, in the following chapter we focus on investigating the potential of CT imaging in VA risk stratification.

3 Fully automatic processing of the cardiac CT imaging

Contents

In this chapter, we propose an automatic image processing pipeline to create a compatible dataset for training a deep learning VA classification model from the cardiac CT images. Basing on the fundamental hypothesis that the VA correlated with the myocardium scar properties, the dataset should implicitly contain features of myocardium scar. The pipeline must be fully automatic in order to limit the human biases and to ensure the reproducibility of the VA classification. The automatic pipeline should limit the time and resources in processing each case, facilitating the inclusion of large number of data in our study.

The proposed automatic pipeline is fed with raw CT images as input and produces a bullseye representation of the LV thickness map. Figure 3.1 shows each processing steps in the pipeline, as detailed in section 3.2. The uniform bullseye map eliminates the 3D heterogeneity of the CT scan LV, while retaining the 3D information in the form of the thickness values. Moreover, the 2D map also limits the calculation resource of the DL model, as compared to the 3D input. The classification model should be able to learn from the heterogeneity of the LV thickness and its relation to the presence of the VA.

Our main contributions in this chapter are:

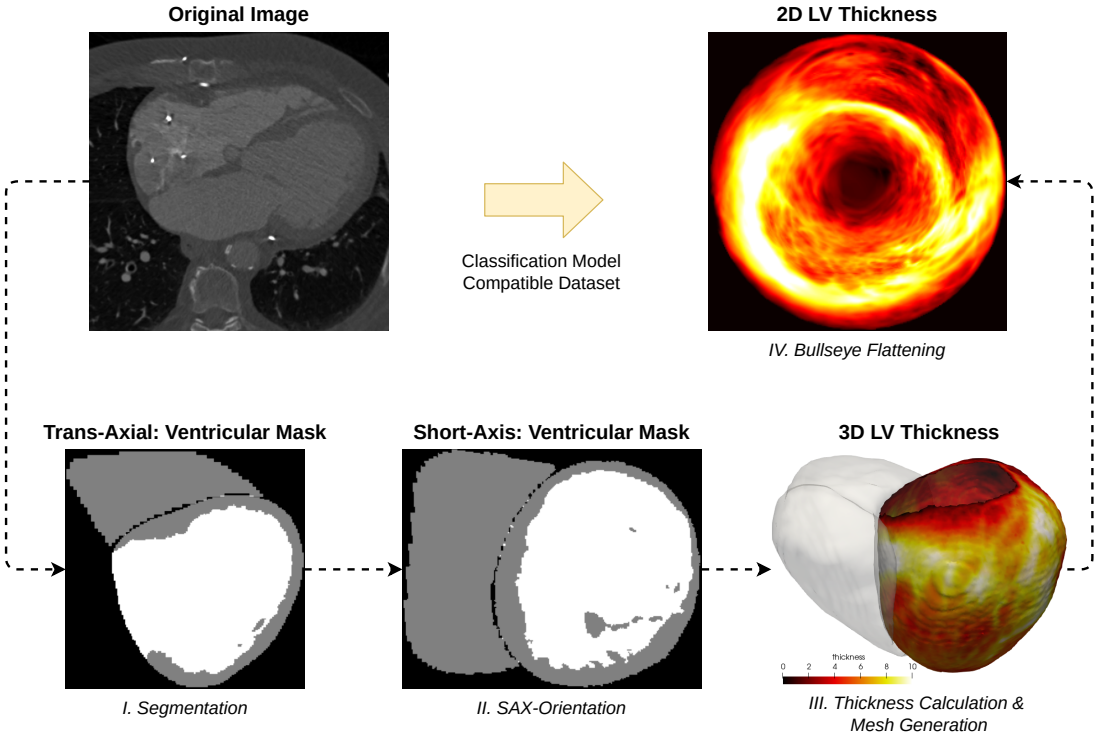


Fig. 3.1.: Automatic cardiac CT processing pipeline to generate the LV thickness bullseye map. The steps include: ventricles (left and right) segmentation, short-axis reorientation, LV thickness value calculation & surface mesh generation and LV bullseye map flattening.

- We proposed a fast and fully automatic pipeline to extract the 3D and bullseye representation of the LV thickness map from 3D cardiac CT images.
- We proposed an automatic parcellation method for the American Heart Association 17-segment model of the LV.
- We integrated an automatic method to select the flattening reference points, ensuring the rotation consistency between the outputs bullseye map.
- Our 2D flattening process allowed the generation of DL-optimal dataset, which both reduces the calculation complexity of the DL model and preserved the scar region position of the 3D images.

3.1 Introduction

Myocardial scar characterisation using late gadolinium-enhanced (LGE) cardiac magnetic resonance imaging is routinely used in the clinics as an evaluation method to improve ventricular arrhythmia (VA) risk stratification. Using this imaging technique, several scar features were found to relate to arrhythmogenicity, including: the scar size; shape; and,

heterogeneity [Kwon, 2009; Heidary, 2010; Zegard, 2021]. However, the generalisation of these findings and their implementation in clinical practice have been often hampered by multiple constraints, such as the availability and reproducibility of CMR acquisition and the robustness of CMR image segmentation methods [McAlindon, 2015]. Moreover, the typical in-plane spatial resolution of CMR is still limited to values > 1mm whereas the slice thickness is 8-10mm, leading to partial volume effects and inaccurate assessment of scar in voxels containing heterogeneous tissue. This is particularly critical in areas of severe wall thinning and at the interface between scar and remote myocardium, regions that are known to have important implication in arrhythmogenicity.

On the other hand, cardiac computed tomography (CT) is more accessible and has been shown to have superior reproducibility and high spatial resolution (i.e., 0.5mm). Although contrast-enhanced CT produces a lower contrast between the scar and remote myocardium compared to late-enhanced CMR, arterial-enhanced CT images give access to a more detailed assessment of LV wall boundaries, facilitating robust and fully automated myocardial segmentation [Jun Guo, 2020]. This allows the assessment of wall thickness (WT) heterogeneity [Cedilnik, 2018], which was shown to correlate with the electrophysiological substrate of post-MI ventricular tachycardia (VT) [Komatsu, 2013]. Several independent studies even suggested that WT heterogeneity may be used to identify critical VT isthmuses [Takigawa, 2019; Ghannam, 2018]. However, this WT-based approach was only employed to facilitate catheter ablation in patients already presenting with arrhythmia; however, its ability to stratify the risk of VA after MI has never been explored.

Our aim is to study the relationship between CT-based WT and post-infarction VA from a large retrospective dataset, using a fully automated deep learning.

3.1.1 Clinical database

Clinical records at Bordeaux University Hospital were thoroughly searched over a period ranging from January 2010 to December 2020, in order to retrospectively identify patients with history of MI who subsequently underwent cardiac CT. The inclusion criteria were: a history of MI before the CT study; an arterial-enhanced and ECG-gated cardiac CT reconstructed at mid diastole; and, the availability of the following clinical characteristics (i.e., age, sex, cardiovascular risk factors, LVEF and the presence/absence of sustained VA). Patients with unknown scar age showing chronic post-infarction scar on imaging and no recent clinical symptoms suggestive of acute coronary syndrome were still considered for inclusion. To avoid confounding factors patients with prior history of LV catheter ablation or surgery, or with a positive diagnosis of hypertrophic cardiomyopathy were excluded. Aortic valve disease and other causes of adaptive hypertrophy were considered for inclusion, given the high prevalence of these co-morbidities in the population of

interest. The information on cardiovascular risk factors and LVEF had to be contemporary with the CT study (6 months without intercurrent event). Regarding VA episodes, only those occurring later than 6 weeks after the MI and prior to the CT scan were retained. These included any episode of sustained VT (>30s), ventricular fibrillation (VF) or aborted sudden death. Taking into account the exclusion criteria as described, the total study population in our dataset was 600 patients, among which 27.5% classified as presence with VA.

All patients provided written informed consent for the secondary use of their clinical data in the present research project, which conformed to the French national law (MR-004) and recommendations from the Commission Nationale de l'Informatique et des Libertés.

3.1.2 Image Acquisition

Over the inclusion period, two dual source CT systems were used (Siemens Somatom Definition and Siemens Somatom Force, Siemens Healthineers, Forchheim, Germany). Images were acquired at the arterial phase of an intravenous bolus of iodinated contrast agent, administered at a minimum flow rate of 5ml/s. Acquisition triggering was achieved with bolus tracking or test-bolus methods, aiming at optimal enhancement within the ascending aorta. The acquisition protocol employed prospective ECG gating in most cases, with different tube voltages and currents depending on the CT system and patient morphology. Images were reconstructed at mid-diastole using an intermediate convolution filter. Note that iterative reconstruction algorithms (applied at intermediate dose) were systematically applied for the most recent studies, but these were not available before 2015. The reconstructed slice thickness varied between 0.6 and 0.7mm, with a field of view close to the heart.

The Fig 3.2 shows a typical example of CT images in our dataset. There are little to no visible differences between the image resolution between the 3 axes. The image contrast also shows clear delineation of the endocardium and epicardium. The presence of the metal artefact caused by the pacemaker, as shown by the red arrows, would have been be a contraindication for this patient from undergoing an MRI scan.

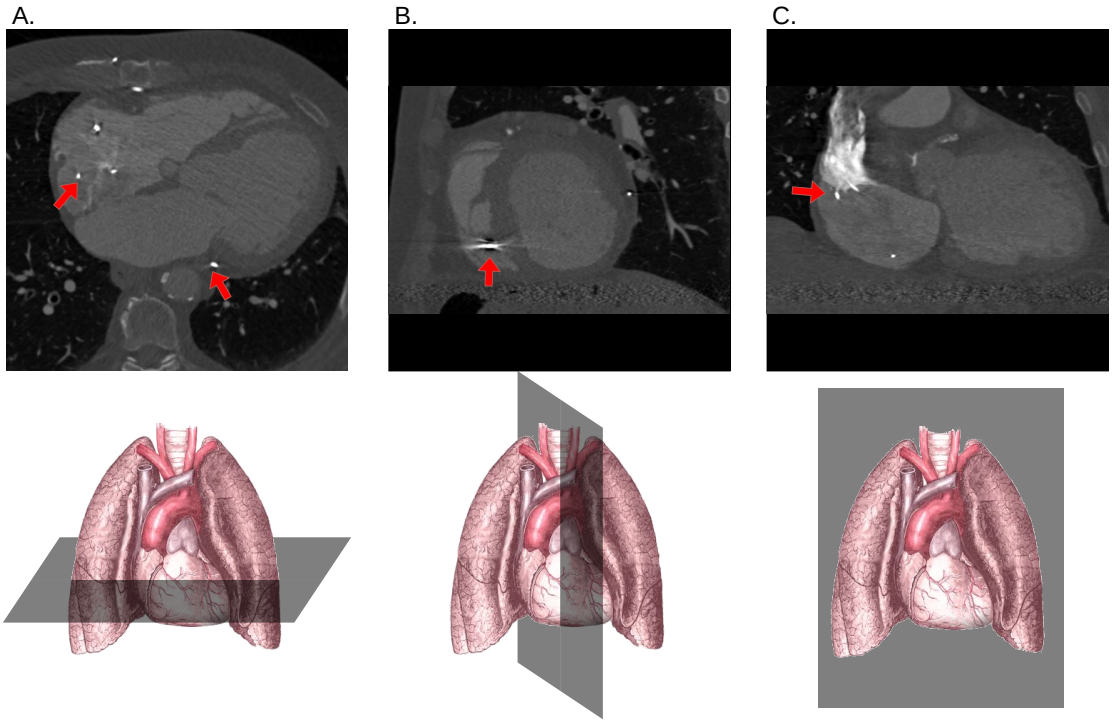


Fig. 3.2.: Cardiac CT imaging. A. Axial plane; B. Sagittal plane; C. Coronal plane. The red arrows indicate the metal artefacts caused by the pacemaker leads.

3.2 Automatic Image Processing Pipeline

3.2.1 Image Segmentation

In order to segment the LV in the CT images, we used the pre-trained segmentation model with the Dual-UNet architecture, as described in the chapter 2. This segmentation network was trained on a separate database of 500 CT images with expert segmentation, in the context of a prior project [Cedilnik, 2018]. The segmentation accuracy of the network was excellent, both for the LV wall segmentation (0.96 Dice score as compared to expert segmentation) and for the WT measurement (95% limits of agreement from -1.28mm to $+0.91$mm). The architecture of the segmentation network is illustrated in figure 3.4.

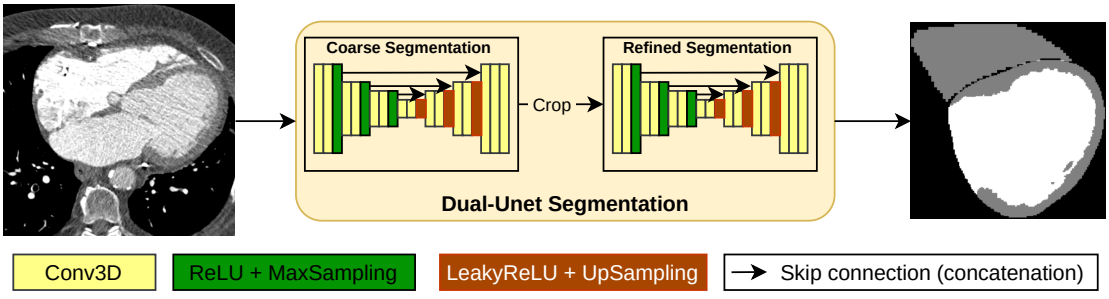


Fig. 3.3.: Dual-UNet Segmentation architecture.

The multi-channel segmentation network was applied directly to our dataset without any additional transfer fitting, to obtain the segmentation of the LV endocardium, epicardium and the RV epicardium. Subsequently, the LV myocardium could be extracted during the post-processing step, and clearly distinguished from the LV epicardium and endocardium contours. The additional post-processing step only includes largest connect component extraction, to remove small detached noises from the output masks.

The outputs from the pre-trained segmentation model demonstrate excellent LV wall segmentation results on the new dataset, as shown in figure 3.4. Our model can accurately segment the thinning regions of the LV wall, as shown by the yellow arrow in figure 3.4.B. Moreover, the model is also robust in the presence of metal artefacts, as highlighted by the red arrow in figure 3.4.A.

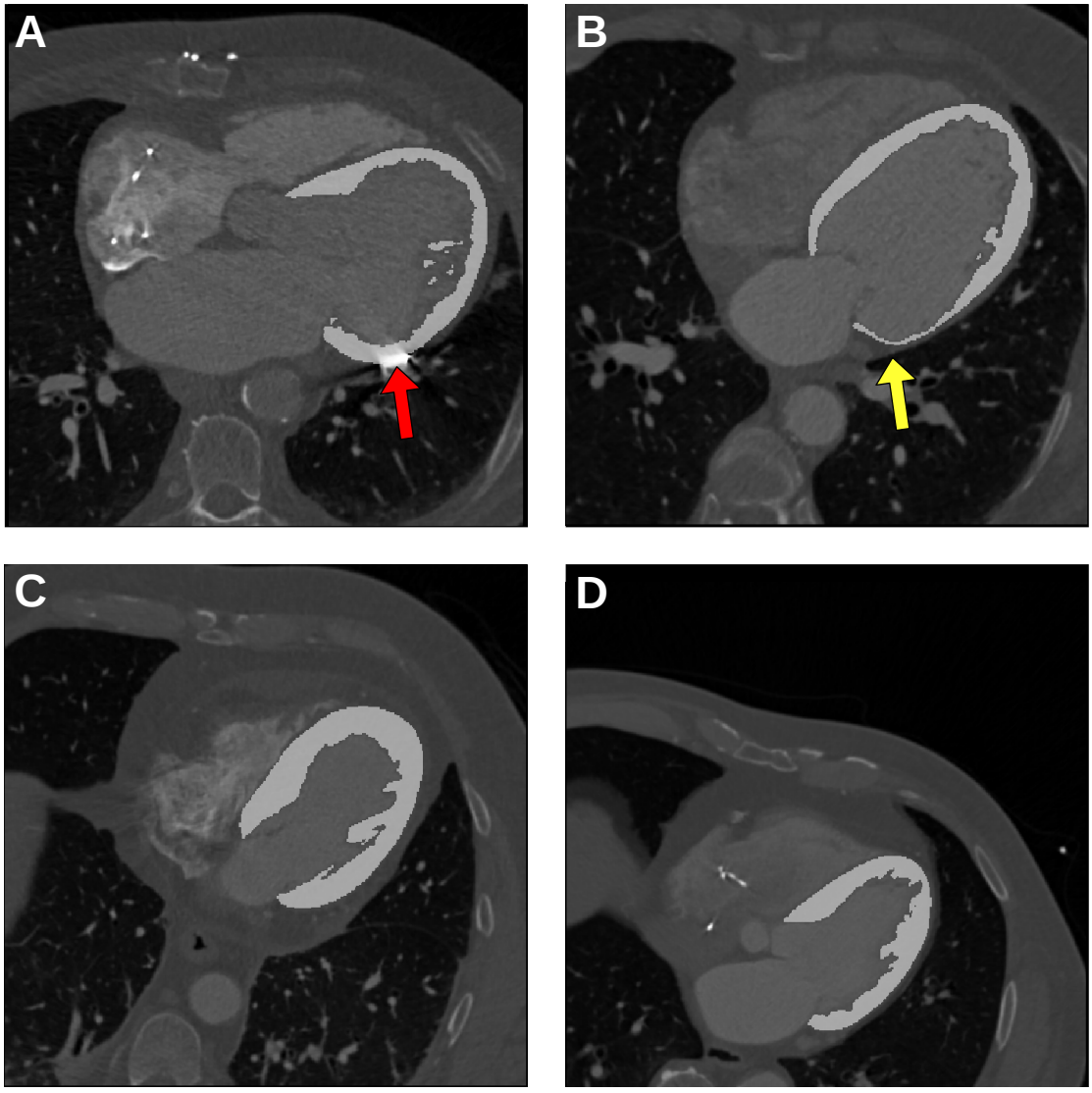


Fig. 3.4.: Dual-UNet Segmentation architecture. Red arrow: metal artefact; Yellow arrow: LV wall thinning region.

3.2.2 Short-Axis reorientation

The raw CT images in our dataset are in trans-axial plane, where the image is oriented with respect to the major axes of the body. However, in the case of cardiac pathologies, the common evaluation of the anatomy and function of the heart is done in the short-axis plane, where the image is oriented with respect to the major-axis of the heart, as referred to the basal and apical axis of the LV. To standardise our dataset and to allow us perform subsequent image processing, in section 3.2.4, we applied an automatic method as proposed in [Nuñez-Garcia, 2020b] to transform the segmented LV masks in the SAX orientation.

The method applied a rigid transformation on the trans-axial images, based on three automatically detected reference landmarks: the LV apex; the mitral valve plane; and, the right ventricle RV insertion in the septum. The final SAX rotation matrix, $R_{SAX} \in R^{3\times3}$, is calculated as the combination of three consecutive rotational matrices: the mitral valve plane alignment R_{MV}, LV septum alignment R_S and the LV long-axis alignment R_{LAX}, as shown in eq. 3.1.

$$R_{SAX} = R_{LAX} \cdot R_S \cdot R_{MV} \tag{3.1}$$

The each rotation matrices of each calculated based on the alignment from the trans-axial unit vector, u_1, and the standard SAX unit vector u_2 of each landmark. The unit vectors u_1 and u_2 of the three alignments were defined as follows:

- R_{MV}: The u_1 is defined as the unit vector of the mitral valve normal and is aligned to the reverse of the z-axis unit vector $\vec{k}$ ($u_2 = -\vec{k} = (0, 0, -1)$).

- R_S: The unit vector u_1 is defined as the horizontal vector pointing from the point on the mitral valve closest to the RV to the centre of the mitral valve. This vector is then aligned with the x-axis unit vector $\vec{i}$ ($u_2 = (1, 0, 0)$).

- R_{LAX}: The u_1 is defined as the unit vector pointing from the apex to the centre of mitral valve and is aligned with the z-axis unit vector $\vec{k}$ ($u_2 = (0, 0, 1)$). The final alignment is done to straighten the alignment of the basal-apical axis of the LV, as the mitral valve plane can be tilted from the segmentation.

During the rigid transformation, we resampled all the images in our dataset using the same reference image with the image origin at $(0, 0, 0)$, image direction as the identity matrix, and the image isotropic resolution of 0.5mm. Figure 3.5 shows examples of the output for the SAX orientation step.

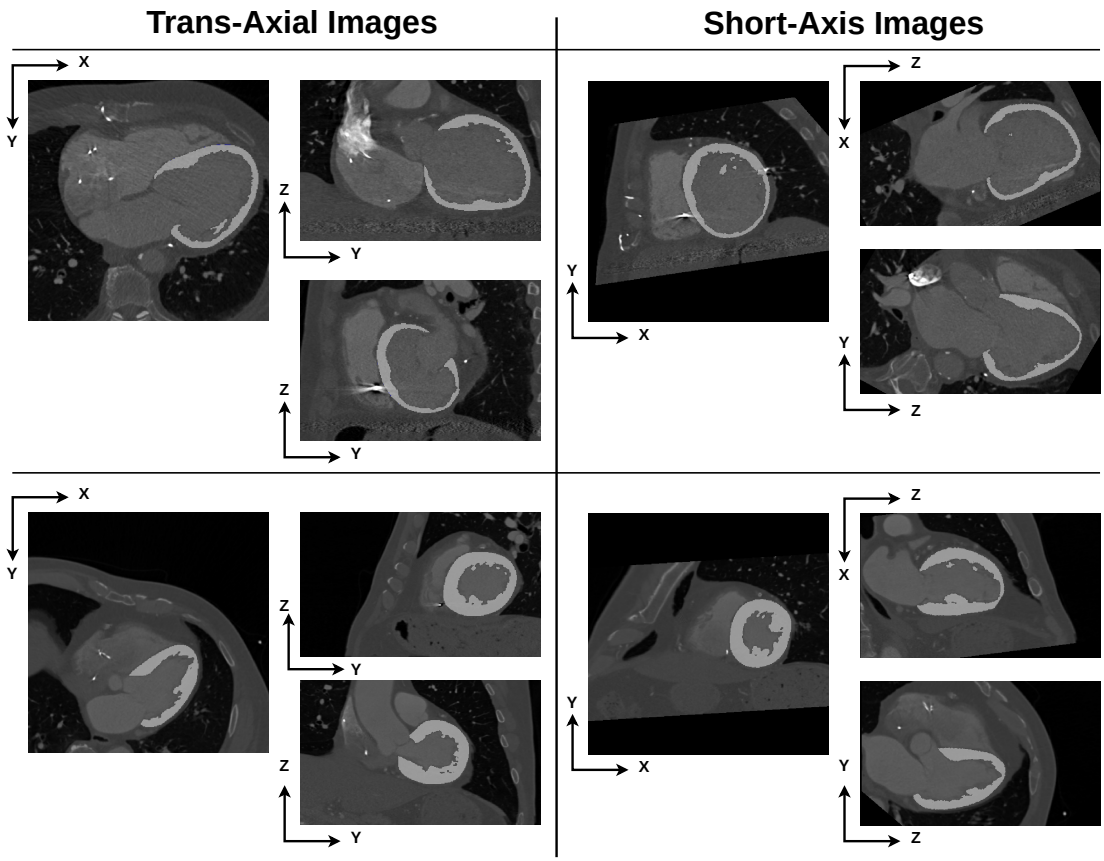


Fig. 3.5.: Comparison between the trans-axial images and the SAX oriented images.

3.2.3 Thickness Calculation & Midwall Mesh Generation

From the LV wall segmentation, WT was calculated as the distance from the endocardium to the epicardium, using a method based on a partial differential equation (PDE) [Yezzi, 2003]. The method defines the tissue thickness as the total arclength of a unique curve, passing by each point x within the tissue, that connects the inner and outer boundary. Following the proposed method, the wall thickness W of each point x on the LV wall was calculated using the following equation:

$$W(x) = L_0(x) + L_1(x) \tag{3.2}$$

where $L_0(x)$ denotes the arclength between x and the inner boundary $\partial_0 R$, and $L_1(x)$ denotes the arclength between x and the outer boundary $\partial_1 R$. The method calculates for L_0 and L_1 by solving the first-order linear PDE conditions in Eq. 3.3 and 3.4, respectively. $\vec{T}$ denotes the tangent field between $\partial_0 R$ and $\partial_1 R$.

$$\nabla L_0 \cdot \vec{T}, \quad \text{with } L_0(\partial_0 R) = 0 \tag{3.3}$$

$$-\nabla L_1 \cdot \vec{T}, \quad \text{with } L_1(\partial_1 R) = 0 \tag{3.4}$$

We calculated the LV thickness values by employing a python implementation of the method made available by other researchers at `https://pypi.org/project/pyezzi/`. The LV epicardium and endocardium mask were used to define the inner and outer boundary. The papillary and trabecular muscle were included as the myocardium for the thickness calculation. We also extracted the arclength L_0 to define the mid-distance between the LV endocardium and epicardium. We normalised L_0 by dividing it to the thickness values. This outputs a LV wall depth map, which assigns the value scale of 0 to 1 for the voxels from the endocardium (inner boundary) to epicardium (outer boundary). We obtained the mid-wall mask by thresholding the wall depth map at 0.5. Figure 3.6 shows the LV wall thickness and wall depth calculation outputs.

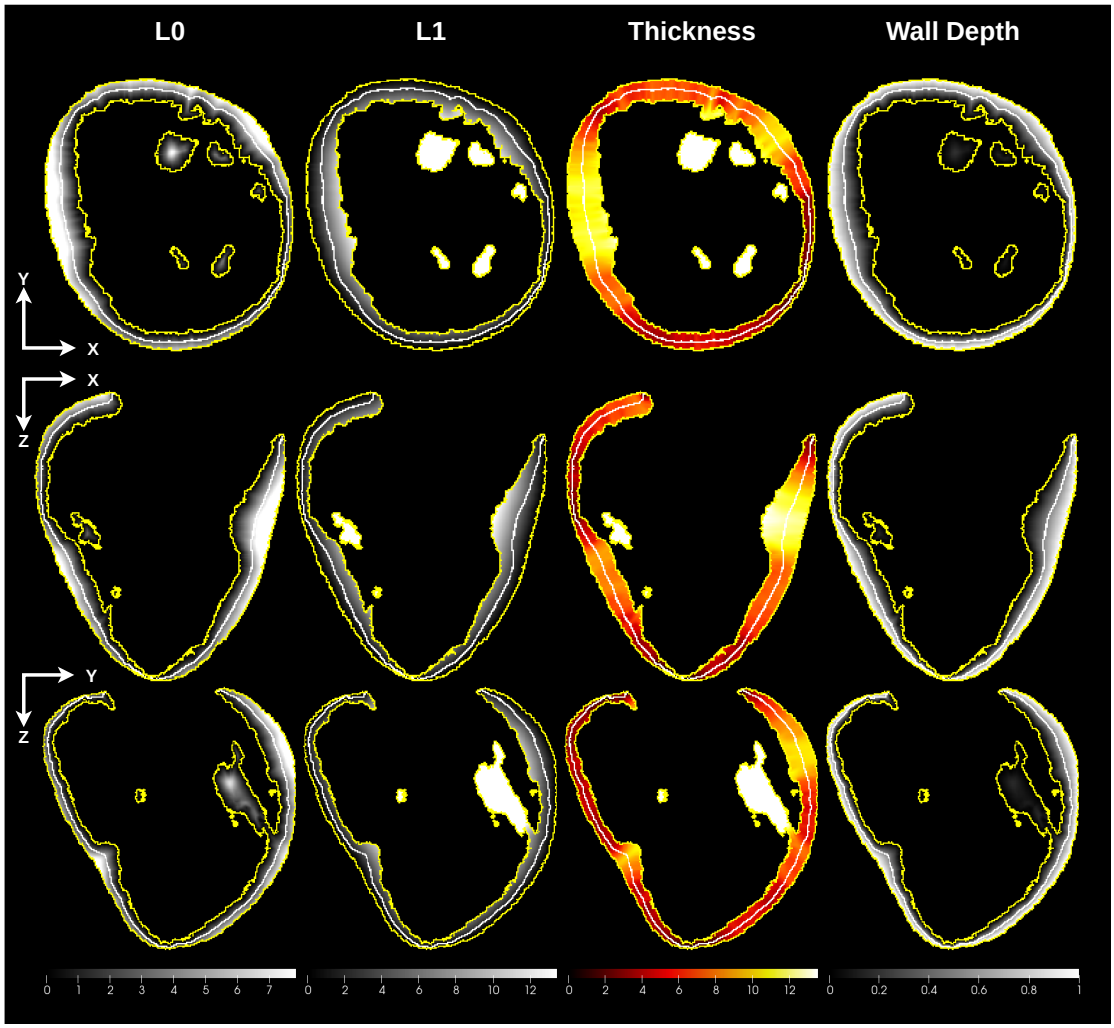


Fig. 3.6.: LV wall thickness and wall depth calculation. The yellow contour marks the LV wall, and the white curve marks the mid-wall boundary of the LV wall.

Finally, we generated the LV mid-wall surface mesh using marching cubes algorithm [Lorensen, 1987] and uniformly remeshed using approximated centroid voronoi diagrams (ACVD) method[Valette, 2008], with the thickness value projected onto the mesh. Figure 3.7 shows an example of the generated LV mid-wall mesh. On the LV mid-wall mesh, the scar size was defined as the surface area with thickness value < 5 mm [Stolzmann, 2008].

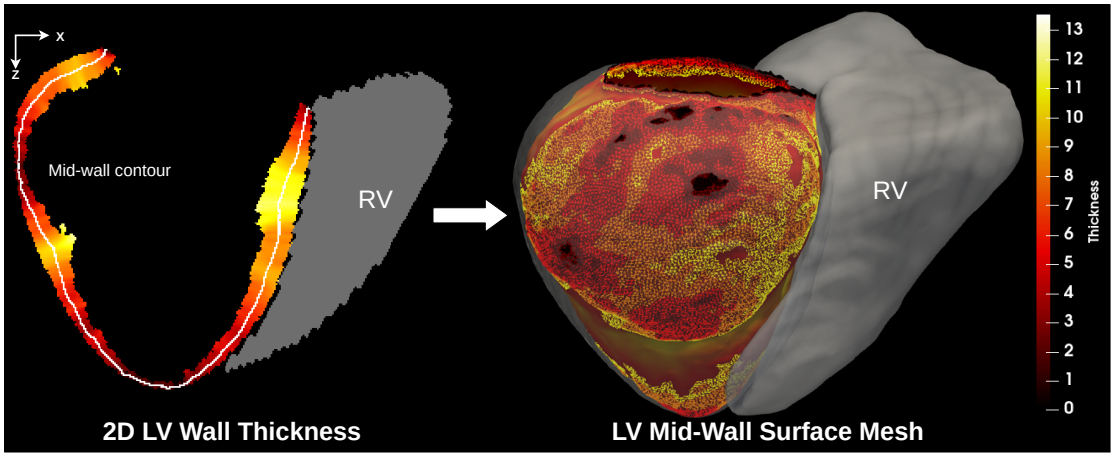


Fig. 3.7.: LV mid-wall mesh generation and thickness value mapping. The wireframe mesh denotes the scar regions thresholded at < 5mm.

3.2.4 LV Bullseye Flattening

Once reoriented, the 3D surface mesh of each patient was transformed into a standardised 2D disk named bullseye map using the method proposed by [Nuñez-Garcia, 2020a], which is based on a quasi-conformational mapping algorithm [Lévy, 2002].

In order to optimise the spatial correspondence between bullseye maps and reduce rotation errors, the method is based on the flattening of multiple landmark reference points onto the original 3D image. Originally developed for LA flattening, the work presented by [Nuñez-Garcia, 2020a] also proposed an LV extension of the algorithm, which is publicly available at `https://github.com/martanunez/LV_flattening.git`. In the original code, the LV flattening required manual selection of the three reference points: one at the centre of the apex and two at the base of the LV. The reference point at the apex served as the centre of the bullseye.

The first of the two the basal points was selected as the point at the centre of the basal septum. It was then placed at 9 o'clock on the bullseye border and further used as the anchor of the rotation. The second point can be selected as any point in the anti-clockwise direction on LV base closed enough to the first point. This second point was used to control the flattening direction.

For our study, thanks to the SAX orientation, we devised a method to automatically select the reference points. We also modified the algorithm to perform the flattening step with only two reference points: the apex of LV (as the centre of the bullseye) and the junction between the 2nd and 3rd segment of the American Heart Association AHA 17-segment model as being the 9 o'clock anchor of the bullseye. The bullseye flattening reference points and output are shown in figure 3.8.

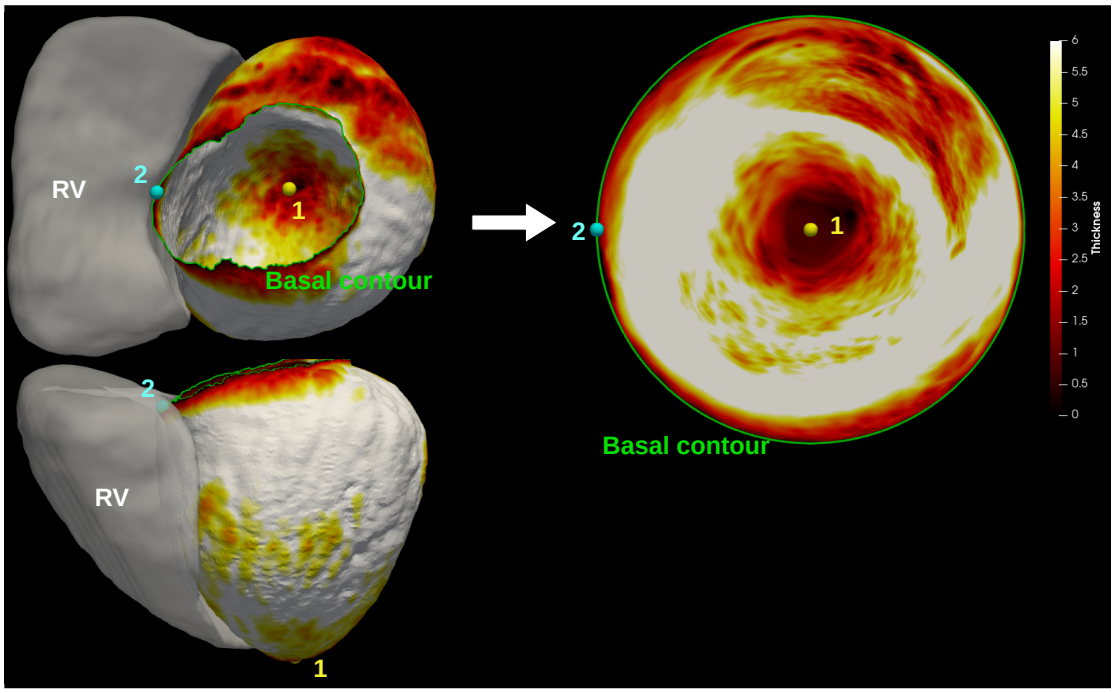


Fig. 3.8.: Automatic 3D to 2D bullseye flattening of the LV surface mesh using quasi-conformal mapping.

3.2.4.1 Automatic parcellation of the American Heart Association 17-segment model of LV

Next, to automate the reference points selection, we proposed an automatic method to parcellate the LV surface mesh into the 17-segment model created by the AHA [Cerqueira, 2002]. The automatic method can be done thanked to the SAX reorientation done in the previous section, which formatted the mesh in the following conditions:

- The principal basal-apex axis followed the z-axis of the euclidean coordinate in the ascending order (the basal region has lower z coordinate than apical region), such that the z-axis can be used for longitudinal parcellation.

- Similarly, the xy plane represents the short axis plane, thus allowing us to perform the circumferential parcellation.

- Finally, the RV is always placed to the right of the LV (in SAX view), which is needed to correctly enumerate the 17 segments.

The 17th segment (apex detection). The detection of the apex is now straightforward since it can be easily detected thanks to the basal-apex axis being along the z-axis. The 17th region was marked as the region toward the apical region started from the first xy plane, with less than 10 endocardium voxels to the end of the LV wall.

Longitudinal Parcellation. The longitudinal parcellation was done by dividing the LV surface mesh between the 17th segment and the lowest z coordinate (i.e., the end of the base) into 3 equal partitions.

Circumferential Parcellation. The circumferential parcellation was done by calculating the angle of each point on the surface mesh. As the parcellation can be done in the xy plan, the angles were calculated as the angular coordinates of the polar coordinate system using the centre of mass of the 17th segment as the centre.

The output parcellation and segment numeration is shown in figure 3.9. The apical reference point was selected as the centre of mass for the 17th segment. The basal reference point was selected as the point on the basal contour at the junction between the 2nd and 3rd segment. During calculation, the basal contour direction was based on the point at the 3rd and 4th segment junction.

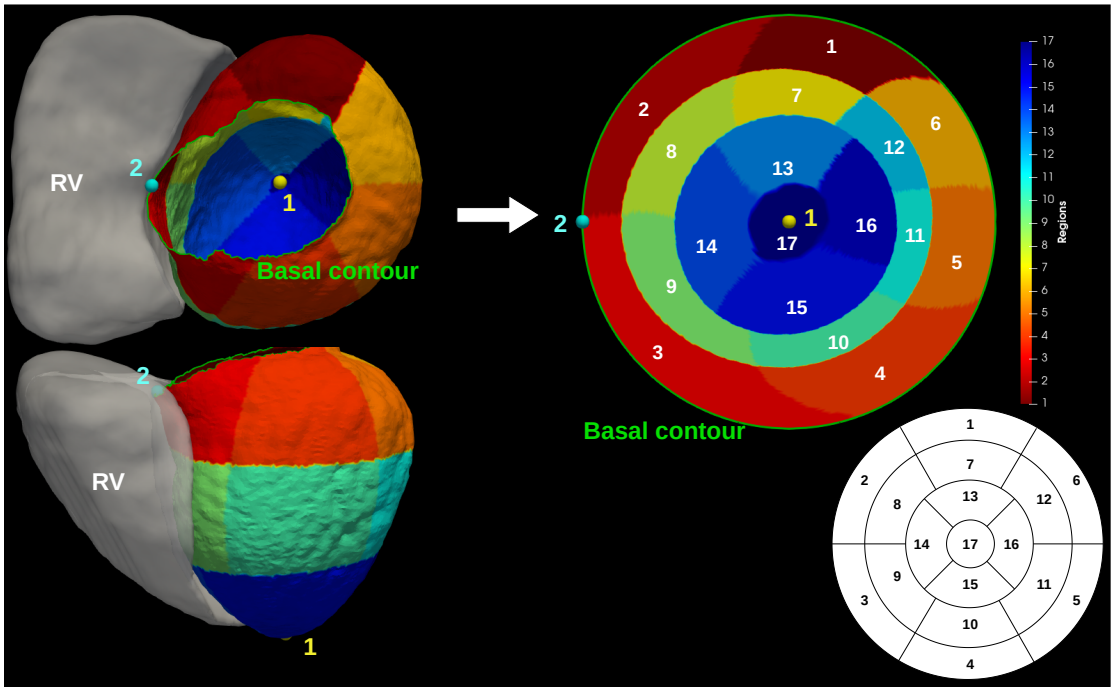


Fig. 3.9.: The 17-segment model. The enumeration (labelling) of the segments was done following the AHA standard convention, as shown in the bottom left corner.

3.2.4.2 DL-compatible Processing

Value Transferring. The (x, y) coordinates of each point of the flatten mesh are normalised between $[-1, 1]$ and centred at $(0, 0)$. To prepare a compatible input dataset for the convolutional DL model, we transform the LV bullseye mesh into a 2D image using the linear interpolation of the corresponding normalised coordinates. First, a 2D blank disk was generated in a desired resolution. Then, the normalised coordinates were calculated for each pixel of the disk following the same conventions. The value of the corresponding coordinates on the bullseye mesh is then calculated using linear interpolation. To generate the dataset, the normalised coordinates of the blank disk can be reused for all the bullseye meshes, thus reducing the processing time.

Using this method, input with different resolutions can be generated. While small resolution image allows a faster calculation (training and inference) of the prediction model, it also reduces the high frequency detail of the original thickness map. In the context of our work, in chapter 4, the resolution of the 2D bullseye images were set to 256×256. The value transferring method could also be used to transfer the value from the 2D image back to the LV bullseye mesh and the 3D LV, as was done for the explainability results in the same chapter.

Thickness Value Normalisation Finally, we normalised the thickness value to between $[0 - 1]$. To ensure the scale consistency of all normalised outputs, we clipped and then divided the thickness map at 10mm. The normalisation of the thickness map X was done using the Eq. 3.5. Note that the value clipping is necessary to control the maximum thickness variety caused by the papillary muscles. Reducing this clipping value also increased the distance between thinning region, as shown in figure 3.10.

$$X^{norm}_{(i,j)} = \frac{max(X_{(i,j)}, 10)}{10} \tag{3.5}$$

3.3 Discussion

Artificial intelligence techniques have already been applied to images for the characterisation of post-MI scar [Chen, 2020]. Most of these works aimed at automating scar segmentation on late gadolinium-enhanced CMR images. However, the robustness of automated scar quantifications on CMR remains limited by the variability of MR sequences [Stirrat, 2015], and often by a low contrast between the subendocardial scar and the adjacent blood pool. Thus, most CMR scar features that were previously proposed to stratify the risk of VA have relied on manual expert segmentation [Kwon, 2009; Heidary, 2010; Zegard, 2021].

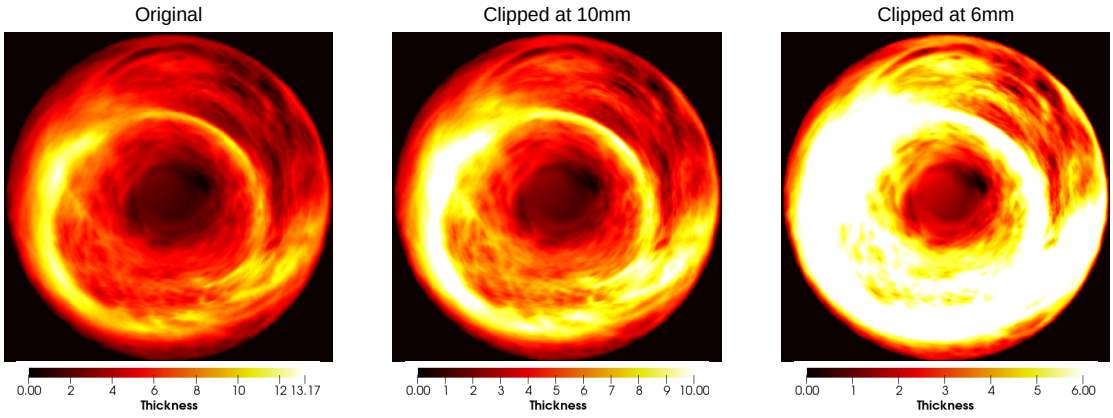


Fig. 3.10.: Thickness value transferring and normalisation. The thickness maps are displayed using normalised colormap at the corresponding scale.

In the present study, the choice of the CT modality was justified by: (i) the broad availability of this modality in clinical practice; (ii) the higher reproducibility of CT image quality across vendors and sites; and (iii) the robustness of fully automated LV wall segmentation on CT images. In addition, we leveraged on prior electrophysiological mapping studies relating WT heterogeneity observed on CT images with the corresponding arrhythmogenic sites found in patients with post-MI scars [Takigawa, 2019].

The favourable qualities of CT images permitted an accurate segmentation of the LV wall by the model trained with a completely different dataset. This eliminates the main bias of our image processing pipeline, the overfitting of the Dual-UNet segmentation model. On top of that, the pre-trained model also demonstrated excellent segmentation accuracy even in the regions presenting pacemaker artefacts, as illustrated in figure 3.4. This further ensured the generalisability of the pipeline against the most limiting artefacts in our target population.

In order to process our dataset, we systematically applied the SAX orientation to all CT images in the dataset. This facilitated the automatic LV flattening step in the current project, while it also made uniform the dataset for subsequent analysis for future projects. The thickness map was then calculated from the transformed LV masks. However, we observed that the nearest neighbour interpolation used during the rigid transformation can generate interpolation artefacts, as shown in figure 3.11. A. The rough contour of the rotated segmentation masks could be corrected by reapplying the segmentation with the SAX oriented CT image. Nonetheless, the effect of the rough contour may be negligible for the thickness value at the LV midwall, as indicated by the normalised histograms in figure 3.11.B, which justified the direct thickness calculation from rotated mask in the current pipeline.

The main goal of the flattening step from 3D to 2D is to construct an optimised dataset, which will be used to train the VA classification model based on the characteristic of

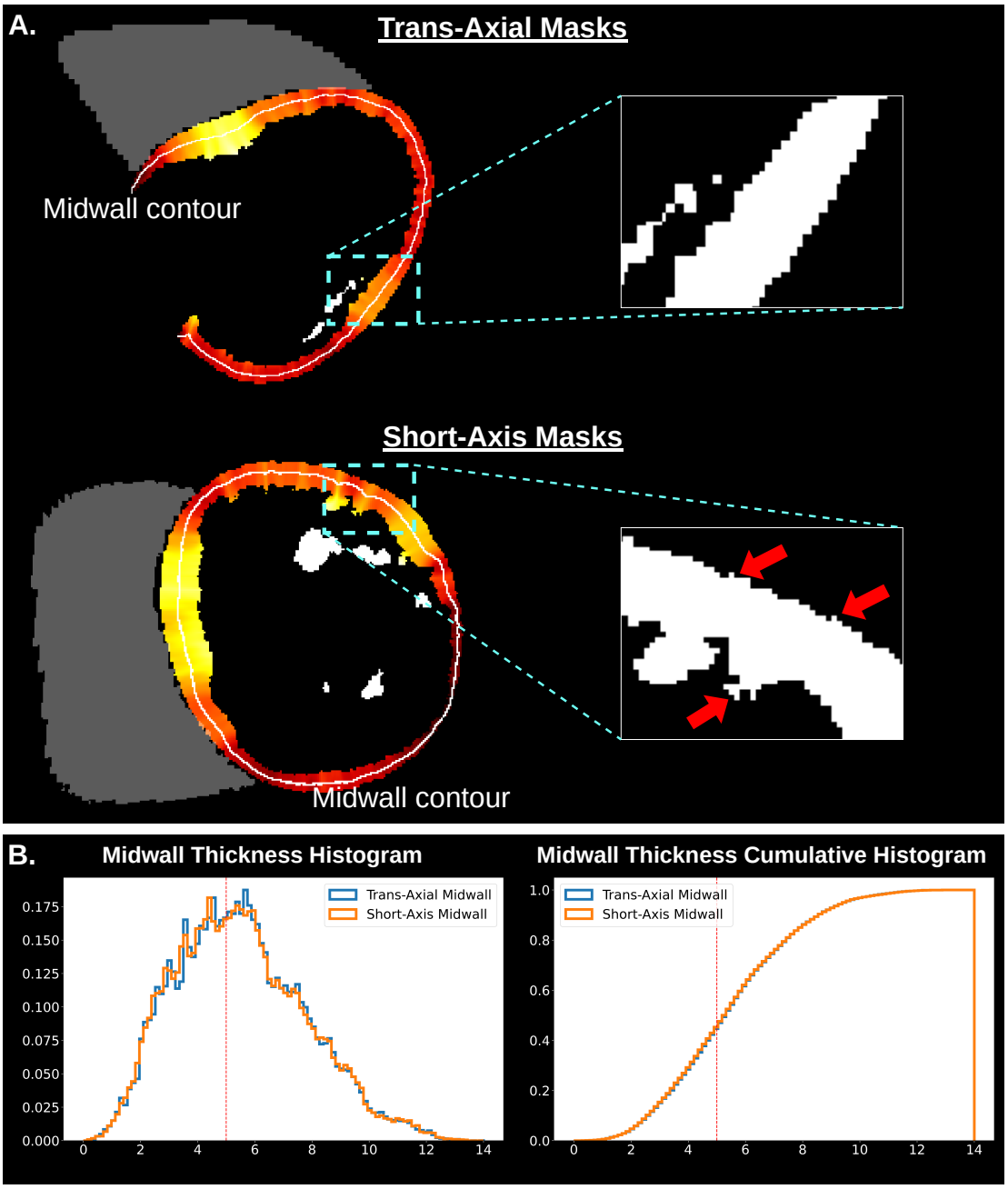


Fig. 3.11.: Comparison between original (trans-axial) and SAX oriented masks for thickness calculation. A. Contour Comparison: notable is the rough contour caused by the nearest neighbour interpolation on the SAX LV wall, highlighted by the red arrows. B. Histogram Comparison: we can observe the histogram consistency between trans-axial and SAX midwall. The cumulative histogram shows a negligible difference between the two midwalls. The red vertical line denotes the 5mm threshold, where the scar size is calculated from the surface mesh.

the thickness value of the LV. The bullseye map eliminates the morphology of the LV from the model optimisation, while the thickness value allows the inclusion of the 3D information. It also removes the blank voxels outside of the LV wall, which accounted for the majority of the input matrix, as shown in 3.12. Moreover, as the rotation is controlled by the reference point, the DL model would be able to fit on the specific position of

the myocardium thinning. This area is important since it would eventually lead to VA generation. Finally, the 2D input also reduces considerably the size of the model since we can use the 2D filters instead of 3D; thus, the total training and inference time will be reduced as well.

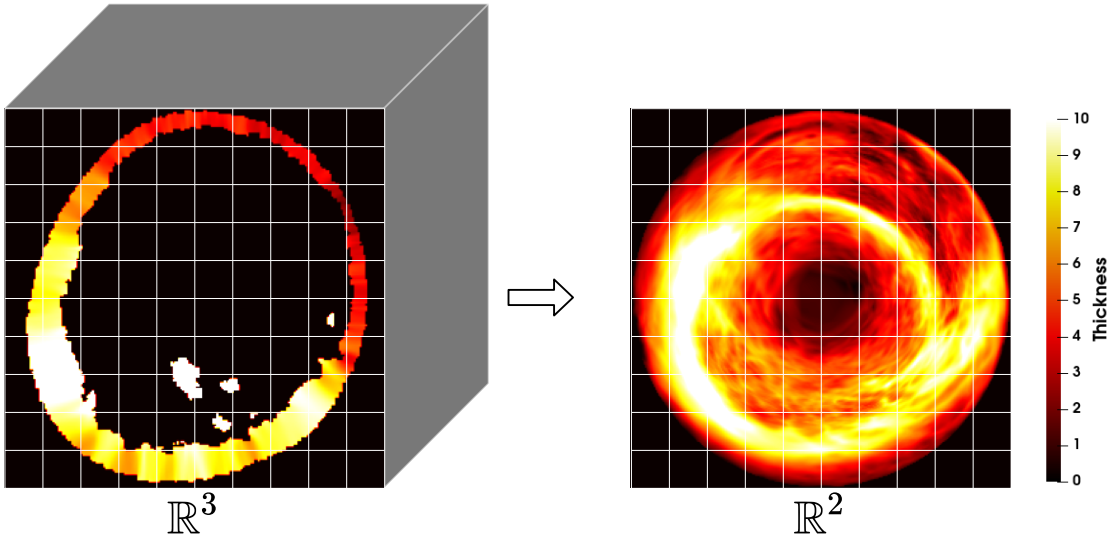


Fig. 3.12.: Comparison between 3D vs. 2D grid input.

Nonetheless, the flattening method introduces non-uniform distortions, with the extended dilation in the apex region, as well as the compression and slice rotation at the basal region, as highlighted in figure 3.13. Fortunately, since similar distortions were applied to all the input data, the DL model would be able to integrate this intrinsic bias during model optimisation, not only in theory but also as proven by our study in the chapter 4.

The total calculation time to complete the process (i.e., from the raw input to the LV bullseye map output) is reported in the Tab. 3.1. The computation was done on a standard performance computer (Inter i7 2.6GHz, x12 CPU and 32 GB RAM). The complete pipeline can be executed in less than 3 minutes.

Tab. 3.1.: Total calculation time.

Processing Steps	**Time (s)**
Segmentation	47
SAX Reorientation	58
Thickness Calculation & Meshing	41
LV Bullseye Flattening	10
Total	156

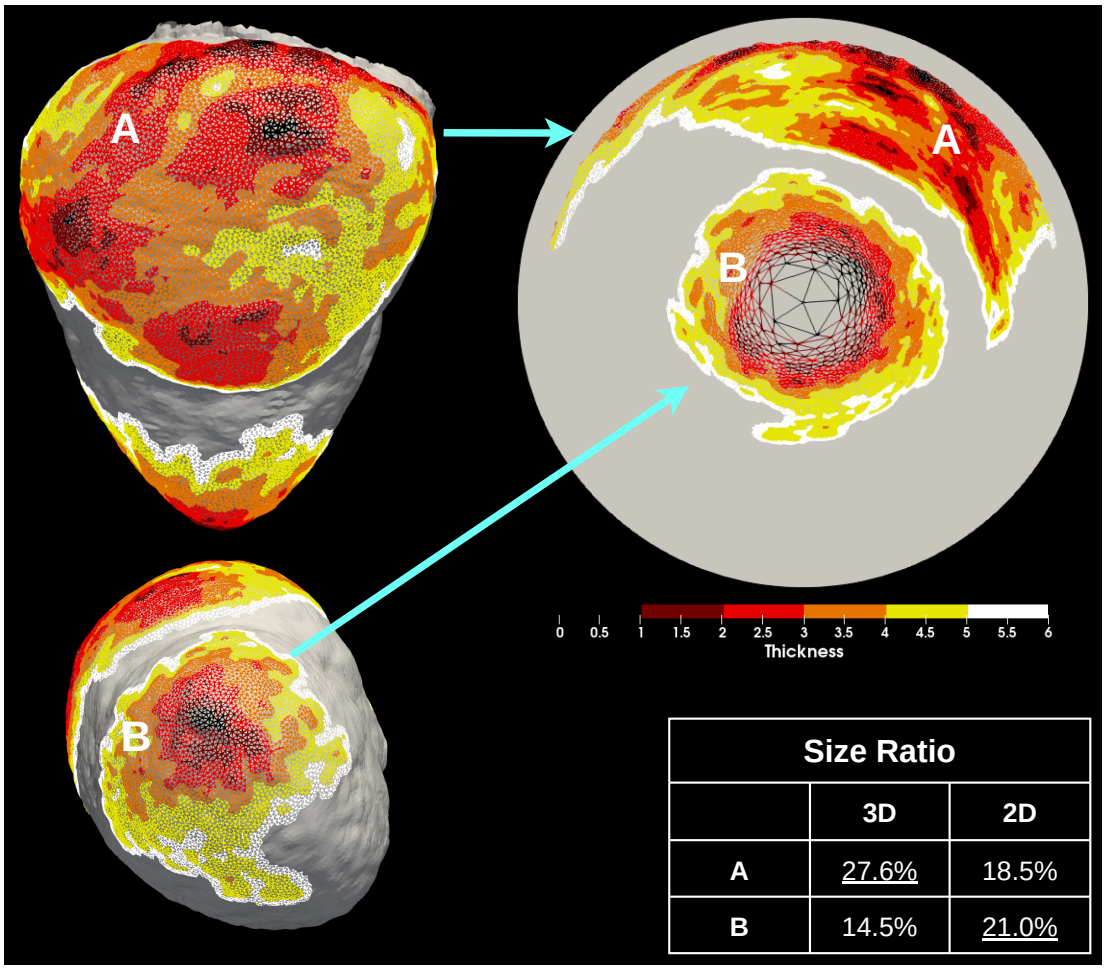


Fig. 3.13.: LV bullseye flattening artefact. The size ratio table shows the basal scar region (A) being compressed by the flattening, and the apical scar region (B). The degree of distortion can be further observed in the different edge length of the wireframe, especially in the apical scar region.

3.4 Conclusion

In this chapter, we proposed a fully automatic and reproducible pipeline to process the raw trans-axial CT image. The main goal of the study was to generate the LV bullseye thickness map dataset that is tailored for the 2D CNN model. Our pipeline can be applied out-of-the-box with no prior transfer learning needed to be adapted to the new dataset, unknown to the segmentation model. To sum up, our multi-step pipeline produces: the trans-axial and SAX segmentation of the LV and RV; the VT-based thickness map of the LV; and, the 3D and 2D LV mid-wall mesh containing the thickness values, as well as the corresponding AHA 17-segment regions.

In the subsequent chapters, we propose different methods to exploit the generated datasets for the prediction of VA using explainable DL models. In the chapter 4, we developed a conditional variational autoencoder model to predict for the VA using the

LV bullseye map. Finally, in the chapter 5, we studied the applicability of the graph convolutional network, which allows us to perform the model fitting and inference directly on the LV mid-wall mesh.

4 Ventricular Arrhythmia Prediction using Explainable Conditional Variation Auto-Encoder Classification

Contents

In this chapter, we propose a conditional variational autoencoder classifier (CVAE-Class) model for the prediction of VA, as detailed in section 4.2.1. To explain the model prediction, we applied the GradCAM++ method, as explained in section 4.2.2. Finally, to study the performance and the generalisability of the classification model, the model was evaluated against the DL baseline model, as well as the clinical risk stratifiers using the bootstrap cross-validation method in section 4.3 and the statistical analysis described in section 4.4.

Using the LV bullseye dataset processed in the previous chapter 3, we explored an automatic method to analyse the LV thickness and its heterogeneity for the VA risk

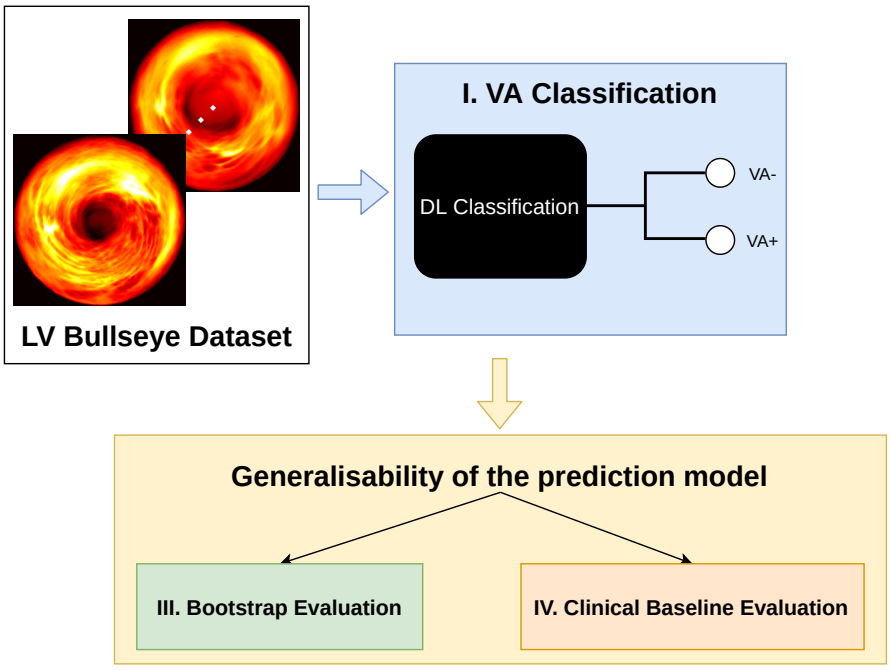


Fig. 4.1.: The chapter road map to build an explainable VA classification model. The DL classification model is detailed in section 4.2.1. We study the explainability of the model in section 4.2.2 and 4.5. We study the model performance and its generalisability using the bootstrap method, in section 4.3, and the statistical analysis, in section 4.4.

prediction. With this respect, figure 4.1 illustrates the summary of the study investigated here.

The results obtained in this chapter proved that the automatic prediction of VA is feasible via the analysis of LV thickness bullseye map using the CVAE-Class model, allowing for a more accurate prediction compared to the clinical markers. We also demonstrate that the model based its VA risk predictions on the CT-defined thinning regions, which solidifies the concept that LV wall thinning plays an important role in arrhythmogenesis.

Our main contributions in this chapter are:

- We proposed a conditional variational auto-encoder classifier model for robust VA risk predictions from the LV thickness bullseye map.
- We proved, after extensive evaluations using the bootstrap cross validation as well as the statistical analysis, that the VA prediction based on the scar size using the CVAE-Class model outperformed the conventional clinical risk stratifiers.
- We also proved that the LV scar size automatically calculated from CT scans can be used as a marker of VA.

- We integrated a post-prediction explainability method, which further validates and justifies that the model predictions are based on the LV wall thinning region.

4.1 Introduction

Advances in Deep Learning (DL) approaches have given rise to several medical image classification networks of exceptional accuracy. Among the different network configurations, here we choose to investigate the embedding-classification architecture where the input image is compressed into its latent representation by the embedding model, after which the latent representation is used as an input to the classification model. While this might appear similar to the ordinary DL classification model, the main difference is the embedding loss used to regulate and train the compression part of the embedding model. This type of architecture has been successfully implemented in classification tasks for several medical imaging applications [Biffi, 2018; Dercksen, 2019; Abbet, 2020]. Multiple DL architectures have been used as the embedding model, such as the deep residual network [Abbet, 2020], the convolutional autoencoder [Dercksen, 2019], and the variational autoencoder [Biffi, 2018]. In our framework, we used the CVAE architecture [Sohn, 2015], since it allows us to robustly reconstruct the input.

The explainable learning is an AI method that explains the model decision according to the original input. An example of explainable learning would be the feature importance techniques, which have been widely studied and adopted in ML. As the usage of DL is being more and more adopted in medical image processing, the explainability of the DL model has evermore become an important branch of DL and AI study. In contrast to typical image segmentation task where the output is already visually justifiable and does not require explanation, the explainable model is crucially needed for diagnosis or outcome prediction tasks. In this chapter, we explored the variant of the gradient class activation mapping method (GradCAM++) proposed by [Chattopadhay, 2018]. The method was chosen as it was proven to be able to more accurately highlight the class specific regions than the previous methods.

The standard risk stratifier for the VA is the LVEF. The common LVEF threshold to classify the patient at risk is at 35%; however, this cut-off value was adopted simply because it was used as one of the inclusion criteria for large scale clinical trails to validate the benefit of ICD therapy [Buxton, 1999]. Therefore, the 35% cut-off value might not be optimal to our dataset. In this study, we implemented different methods to adjust the LVEF value to our dataset both by recalculating the optimal cut-off value as well as by using a regression model. Finally, in order to discover any potential VA biomarkers in our database, we ran a complete statistical analysis on all of the available clinical information.

4.2 Method

Specifically, using the LV bullseye, here we proposed an automatic classification using the conditional variational autoencoder classifier model. At inference, for each prediction, an activation map is generated, using the GradCAM++ method, to explain the model prediction. Figure 4.2 illustrates the fully automatic pipeline for the VA prediction using the CT image input.

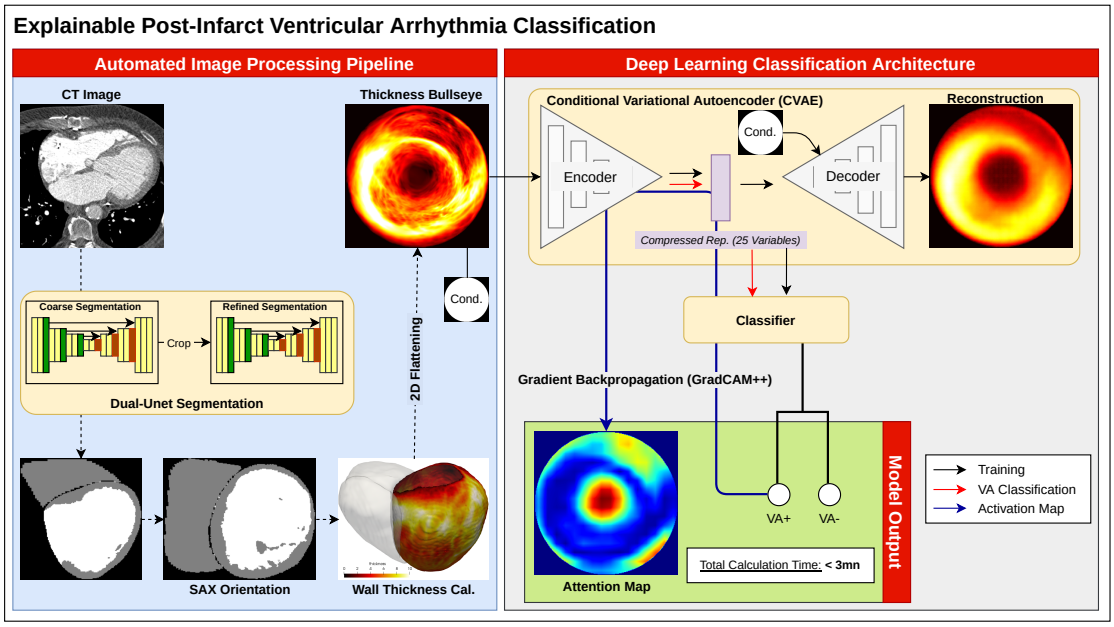


Fig. 4.2.: The complete automatic pipeline for VA using CT imaging.

4.2.1 Conditional Variational Autoencoder Classifier Model

The CVAE-Class network is a combination of a CVAE and a fully connected classifier model. The detailed model architecture including the layer type, the number of filters and loss function used are shown in figure 4.3. For this project, the network was built and trained using the Tensorflow 2.0 package[1].

4.2.1.1 Conditional-VAE Model

Variational Autoencoder (VAE). The VAE model follows the same structure as the autoencoder (AE) model, where the input is compressed to a low-dimensional representation then decoded back to the original data. The AE model is optimised only with the input data, thus giving it the prefix 'auto-'. The VAE introduces the variational inference step (based on the Bayesian distribution) to the AE model, allowing the model to be used in the generative task. Let us suppose that we have the probabilistic encoder

[1] https://www.tensorflow.org/

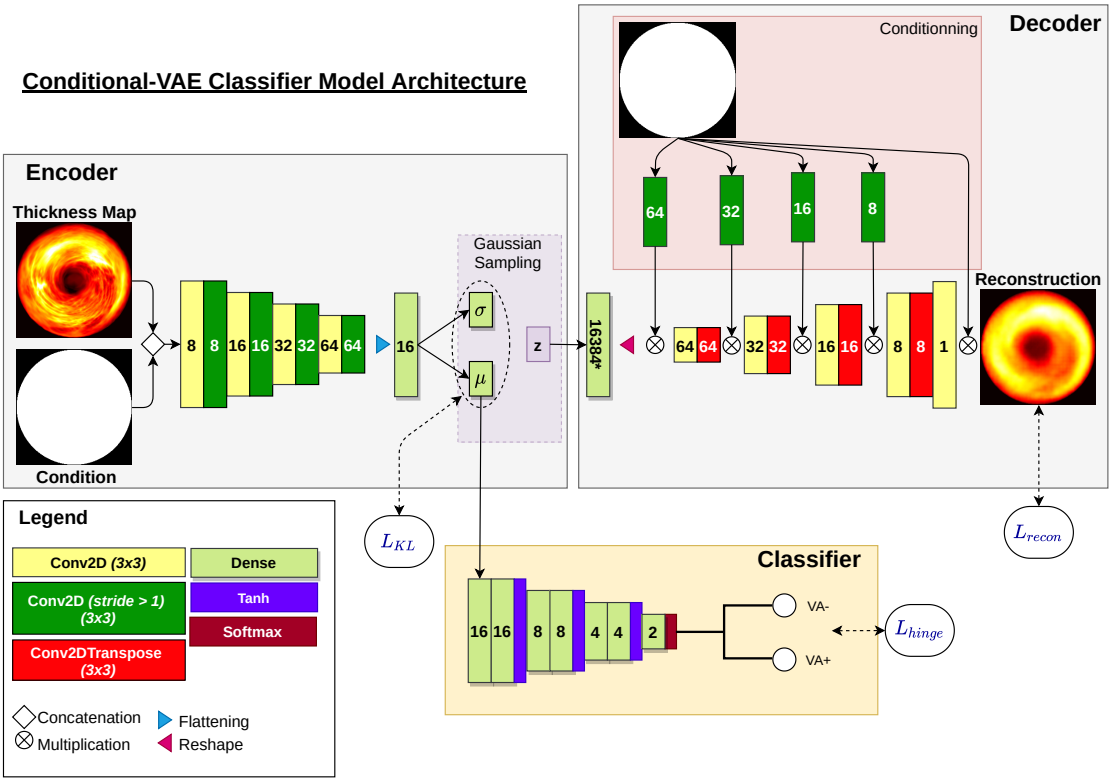


Fig. 4.3.: CVAE-Class network architecture. The number in each layer indicates the number of output filters (convolutional layer) or units (dense layer).

model $q_\phi(z|x)$, which outputs the latent z knowing the input x, and the decoder, or the generative model, $p_\theta(x|z)$. The model is then fitted to optimise the weights parameter ϕ and θ using the following the eq. 4.1.

$$L_{VAE}(\theta, \phi) = -\log p_\theta(x|z) + D_{KL}(q_\phi(z|x)||p(z)) \tag{4.1}$$

The first term corresponds to the reconstruction cost between the original input x and the decoded $x \sim p_\theta(x|z)$. The second term D_{KL} denotes the Kullback-Leibler divergence between the encoder distribution $q_\phi(z|x)$ and the prior distribution $p(z)$, which can be calculated using the following equation:

$$D_{KL}(q_\phi||p_\theta(z)) = \sum q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z)} \tag{4.2}$$

Here the assumption is that both $q_\Phi(z|x)$ and $p(z)$ are Gaussian. The prior over the latent variables is set to an isotropic multivariate Gaussian $p_\Theta(z) = \mathcal{N}(z; 0, I)$. The encoder output vector z can be sampled from the posterior $z \sim q_\Phi(z|x)$ by $z = \mu + \sigma \odot \epsilon$, where μ

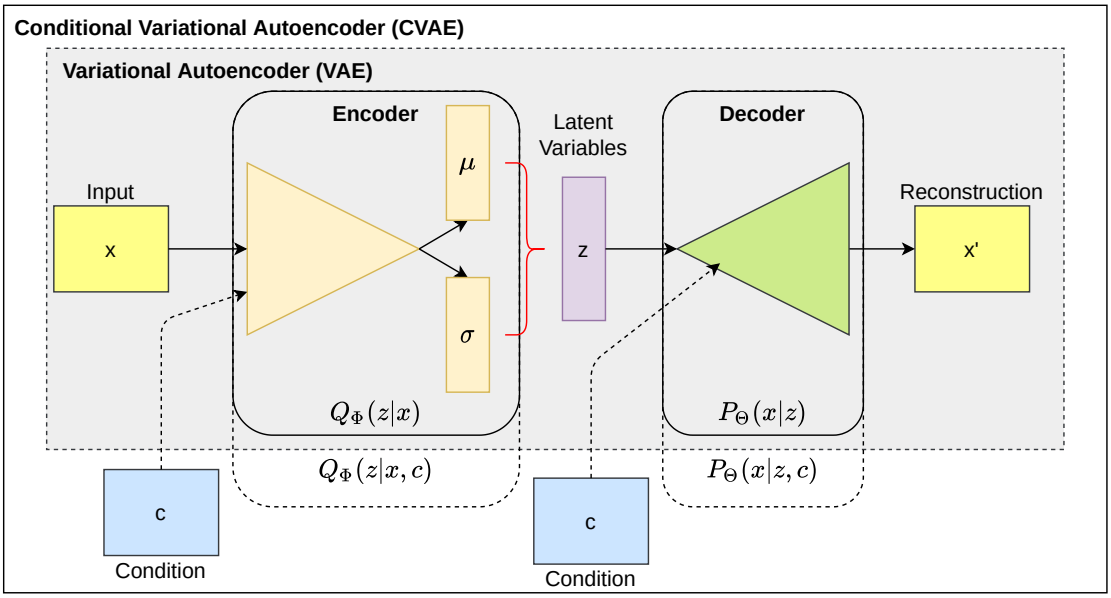


Fig. 4.4.: The schematic representation of the VAE and Conditional-VAE (CVAE) model. The condition input is fed to the CVAE as the input to the encoder and to the decoder.

and σ are the outputs of the encoder signifying the mean and standard deviation, and ϵ denotes the noise variable $\epsilon \sim \mathcal{N}(0, I)$.

Conditional-VAE (CVAE). The conditional variational autoencoder is first proposed by [Kingma, 2014], where the conditional input c is introduced into the encoder and decoder of the VAE. The condition c is used as the prior to control the decoding reconstruction, as well as to improve the reconstruction quality. The encoder and the decoder distribution can thus be written as: $Q_\phi(z|x,c)$ and $P_\theta(z|x,c)$. Figure 4.4 illustrate the schematic representation of the CVAE-type model.

Model Architecture. In the encoder branch, we used a concatenate layer to combine the main input image with the condition mask, followed by a series of 1-stride and 2-stride convolution layers. The 2-stride layers progressively reduced the resolution from 256×256 to 16×16. At the bottleneck, we flattened the output features before applying the fully connected layers to generate the latent mean μ and log variance σ vector. Thereafter, we applied the Gaussian sampling method to generate the latent vector z. The latent dimension was set to 25.

In the decoder branch, we used a fully connected layer in order to increase the number of features of the latent vector z. Then, we reshaped the vector to the same shape as the one prior to the flattening layer. The decoder was then built using the sequences of regular and transpose convolution layers, respectively. Notably, we set the stride of the transpose layer to 2 to increase the resolution of the output. Also, at each decoding level, we combined the condition mask and the main output using a multiplication step. A convolution layer with an adjusted number of stride was used to rectify the resolution

difference. Lastly, we multiplied the features with the conditional mask to get the final reconstruction output. We conditioned this VAE with the mask in order to prevent the network from using its parameters to correctly predict the circular shape of the map.

Loss Functions. The CVAE was trained using the sum of the reconstruction loss (L_{recon}) and the Kullerback-Leibler divergence loss (D_{KL}), eq. 4.3. The L_{recon} was calculated using the weighted root mean square error, defined in eq 4.4, where N is the number of pixels, $\hat{y}$ denotes the decoder reconstruction output, y the input LV bullseye and m is the binary mask. The loss function puts higher weights on the thinning regions compared to the healthy regions. Note that D_{KL} was calculated as the difference between the distribution of the latent variables $\mathcal{N}(\mu, \sigma^2)$ and the normal distribution $\mathcal{N}(0, 1)$. Empirically, the β value was set to 0.01, in order to provide the best reconstruction and classification accuracy.

$$L_{CVAE} = L_{recon} + \beta D_{KL} \tag{4.3}$$

$$L_{recon} = \sqrt{\frac{1}{N}\sum_{i}\sum_{j}(\hat{y}_{ij} - y_{ij})^2 \times (2m_{ij} - y_{ij})} \tag{4.4}$$

4.2.1.2 Classifier Model

The classifier model used the latent mean μ as input. The model was built using three blocks of two fully connected layers and a hyperbolic tangent activation layer, tanh. The numbers of units were set to 16, 8 and 4, respectively. In order to get the classification outputs (healthy/VA- and VA+), we used a fully connected layer with 2 units followed by the softmax function.

The classifier was trained using the hinge loss, as defined in eq. 4.5, where N denotes the number of classes, y denotes the ground truth and $\hat{y}$ denotes the classification score. Note that the expected value of the ground truth y for each class is between [-1, 1], while the classification score $\hat{y}$ is between [0, 1].

$$L_{hinge} = \frac{1}{N}\sum_{i}^{N}\max\{1 - y_i * \hat{y}_i, 0\} \tag{4.5}$$

4.2.2 Model Explainability via GradCAM++

To explain the model prediction, we used the GradCAM++ ('grad-cam-plus-plus') method which is based on the gradient backpropagation of the prediction score to generate the class activation mapping. This method is a derivative of the GradCAM method, which in turns is derived from the original class activation mapping (CAM) method.

4.2.2.1 Class Activation Map: CAM, GradCAM and GradCAM++

The CAM method was proposed by Zhou et al. [Zhou, 2016] and can generate a value map highlighting the regions of the predicted class. For example, the activation map of the "dog" classification highlights the regions on the input image where the dog can be found. Although, it might share some similarity with the segmentation task, the activation map is solely generated from the classification-based model, which does not require any training with the ground truth masks. As such, beside its main used in model interpretation, it has been successfully used in weakly supervised segmentation task [Kang, 2021], to improve the segmentation model [Lin, 2020], as well as other class-specific visualisation tasks such as object localisation and pattern discovery [Zhou, 2016].

The CAM method and its variants follow the same fundamental assumption that the model prediction for each class can be linearly calculated using the eq. 4.6, where y^c denotes the prediction score of class c, A_k denotes the outputs of the last convolutional layer of k filters, while w_k^c is the weights designated for the kth filter.

$$y^c = \sum_k w_k^c \sum_i \sum_j A_{ij}^k \tag{4.6}$$

From there, the activation map M can be defined with eq. 4.7.

$$M_c(i,j) = \sum_k w_k^c A_k(i,j) \tag{4.7}$$

The methods proposed by CAM, GradCAM and GradCAM++ to calculate the weights w_k are illustrated in figure 4.5. The CAM method requires the classification model to have a specific architecture configuration, the global average pooling to be followed by the fully-connected layer and the softmax activation, where the w_k is the weight from the fully-connected layer. The GradCAM method proposes the w_k calculation using gradient backpropagation, thus lifting the architecture limitation of the previous method. Finally,

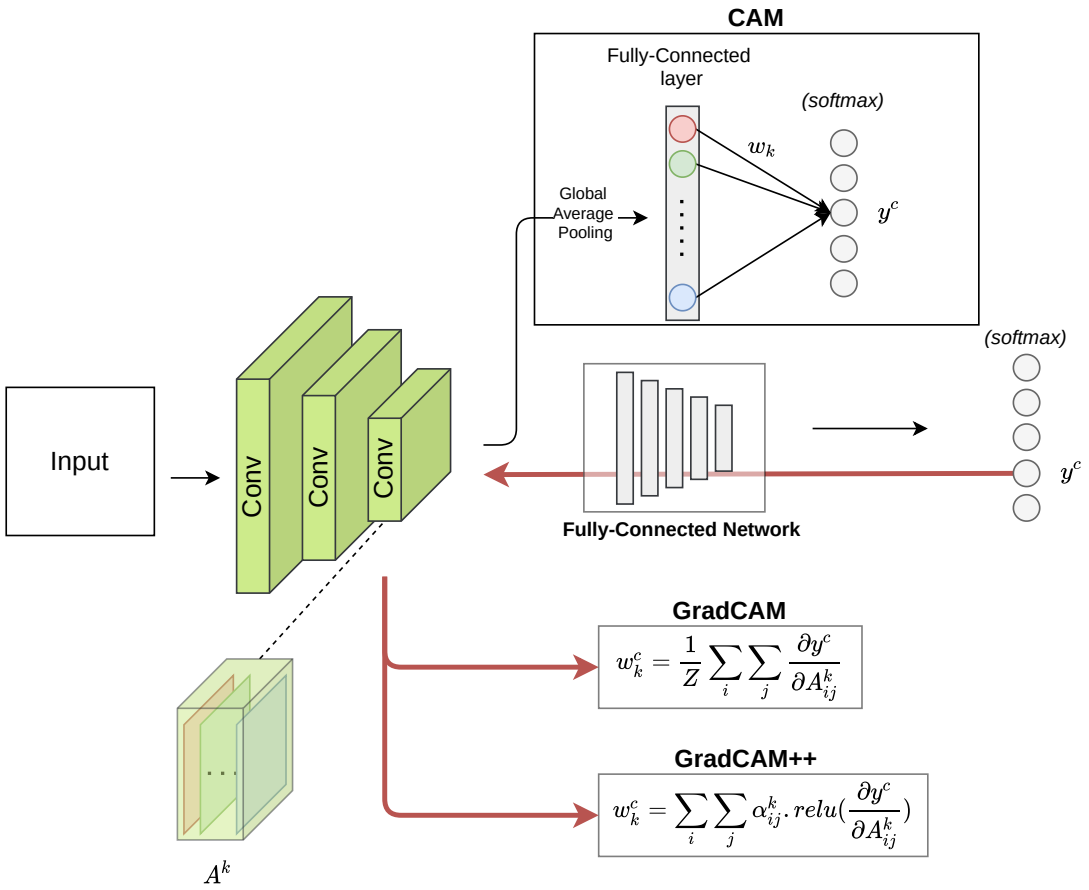


Fig. 4.5.: The comparison between CAM, GradCAM and GradCAM++ calculation.

the GradCAM++ proposes the positive gradients and additional factor α, calculated from the third degree partial derivative, as shown in eq. 4.8. Here (a, b) are the iterations over the same activation map A^k as (i, j).

$$\alpha_{ij}^{kc_1} = \frac{\frac{\partial^2 y^{c_1}}{(\partial A_{ij}^k)^2}}{2\frac{\partial^2 y^{c_1}}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \frac{\partial^3 y^{c_1}}{(\partial A_{ij}^k)^3}} \tag{4.8}$$

We adapted the GradCAM++ method in our pipeline to generate the class activation map to explain the model prediction. As the CVAE-Class model only output two classes (VA- and VA+), we only calculated the activation map for the VA+. This would allow us to verify how the characteristics of the LV bullseye contribute to the VA+ prediction. The schema for activation map calculation for the CVAE-Class model is shown in figure 4.6.

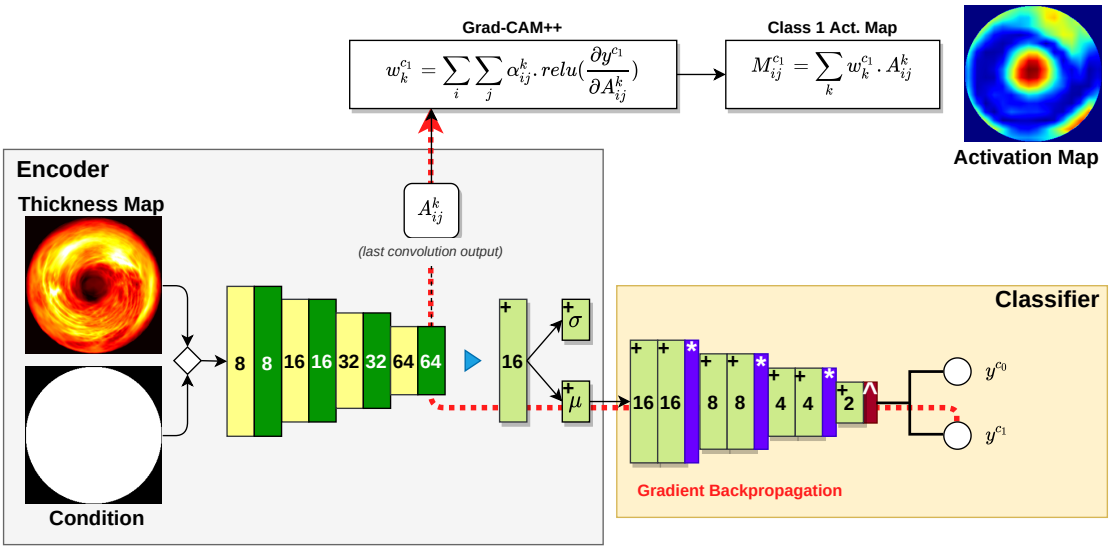


Fig. 4.6.: GradCAM++ calculation scheme. For each prediction, the method backpropagated the gradients from the prediction score through the network to generate an attention map, where the high attention regions (in red) corresponded to regions used by the model to obtain its prediction.

4.3 Model Evaluation with Bootstrapping Method

4.3.1 Experimental Setup

In this section, we evaluated the CVAE-Class performance against the prediction models based on LVEF, as well as a DL baseline model. From the CT dataset described in section 3, we only considered the patient data with the LVEF information, which resulted in a total of 686 patients (mean age 73 ± 12, 83.4% men, and 26.5% with VA) included, in this evaluation.

Next, the dataset was randomly split into the training dataset and the testing dataset, respectively, which were further used to train and to evaluate each prediction model. In order to prevent sampling bias, we applied the bootstrap resampling method proposed by [Efron, 1993] to run the cross-validation of the models.

4.3.1.1 Bootstrap Resampling

The three fundamental concepts of statistical theory employed are: data collection; summary; and, inference. These concepts can be simplified by the following questions:

1. How should the data be collected ?

2. How should the collected data be analysed and summarised ?

3. How accurate are these summarises/conclusions ?

As defined by the author in [Efron, 1993], the bootstrap is a statistical inference derived from the data-based simulation method. The term “bootstrap” is based on the well-known phrase “pull oneself up by one’s bootstrap”, which referred to the iterative resampling, with replacement method used in the bootstrap method. In short, the bootstrap method is proposed to infer about the population from the average of multiple small resamplings of the sample data.

In ML, the central problem in model validation is the bias when testing the dataset, where the data characteristic in the test dataset might favour one model instead of another. The test dataset distribution problem may be solved by the different variant of cross-validation method; however, the bootstrap method allows the reliable confidence interval calculation, which is not readily available with the cross-validation methods.

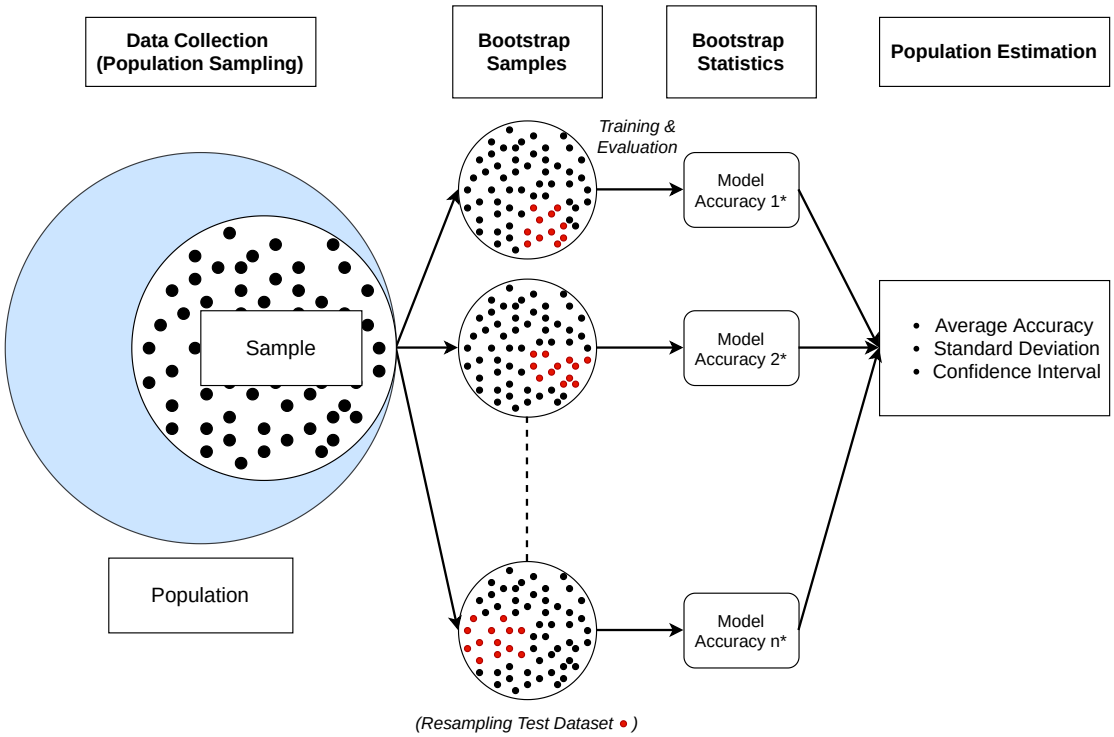


Fig. 4.7.: Bootstraping test dataset. Suppose our dataset was collected from a larger population distribution, and the test dataset was resampled with replacement n times to simulate the sample collection from the population. The model is then trained (with the remaining data) and evaluated on the each test sets. Finally, estimations can be done to infer the model performance on the population.

In this study, we adopted the bootstrap method for the model validation, as shown in figure 4.7, using the following steps:

1. Set the number of bootstrap resampling

2. For each resampling:

    a) Sample with replacement the test dataset

    b) Train the model with the data that was not chosen as the test dataset

    c) Run the evaluation on the test dataset

3. Calculate the average model performance (mean and standard deviation) and the confidence interval.

Confidence Intervals. As the bootstrap sampling simulated the population sampling from a single sample data pool, several methods have been proposed to accurately define the confidence interval of the inference. Such methods include: the percentile

method, bootstrap-*t* interval, the bias-corrected accelerated (BCa), and the approximate bootstrap confidence intervals (ABC). Specifically, in this study we employed the ABC non parametric method to calculate the confidence interval of the average model prediction accuracy of the bootstrap test.

Both BCa and ABC method introduce the bias-correction and acceleration terms into the estimation of the confidence interval. The bias-correction measures the inconsistency between the median of inferences, while the acceleration term measures the rate of inference change as respected to the true value. These allow us to better estimate the confidence interval compared to the simple percentile method. The ABC method is the approximation method for the BCa, which gives satisfactory result in the case of limited number of bootstrap sampling. The method bases its confidence interval approximation on the Tyler expansion series. While the BCa method derives its bias-correction and acceleration term by the second order resampling (resampling of the bootstrap resampling), the ABC method allows the calculation of these terms using numerical second derivatives.

4.3.1.2 Training Setup

To account for the computation time of the deep learning model, the number of resampling was set to 25, which has been shown to be sufficient in estimating the randomness caused by the sampling data [Efron, 2015]. For each resampling, we randomly sampled 100 cases (i.e., 50 healthy subjects and 50 VA patients) to be used as testing dataset. The population sampling diagram is shown in figure 4.8.

CVAE-Class Model. The CVAE-Class networks were trained step-by-step, where we first trained the CVAE to adequately embed the input before optimising the classifier. Both models were optimised using Adam optimiser with the initial learning rate at $1e-4$. The learning rate was reduced by half after 5 epochs with no validation improvement. In the training dataset, in order to balance between healthy and VA patients, we trained and validated the model with the same number of data from each class at each iteration. The training was set to stop when the validation loss does not improve after 50 epochs. The model was re-run 5 times, afterwards the model with the highest validation accuracy was selected.

Baseline Models. To compare the CVAE model configuration, we built a straightforward image classifier network using the architecture of the encoder and the classifier of the CVAE-Class network. The model was trained with the same LV bullseye input and training hyperparameters as the CVAE-Class model.

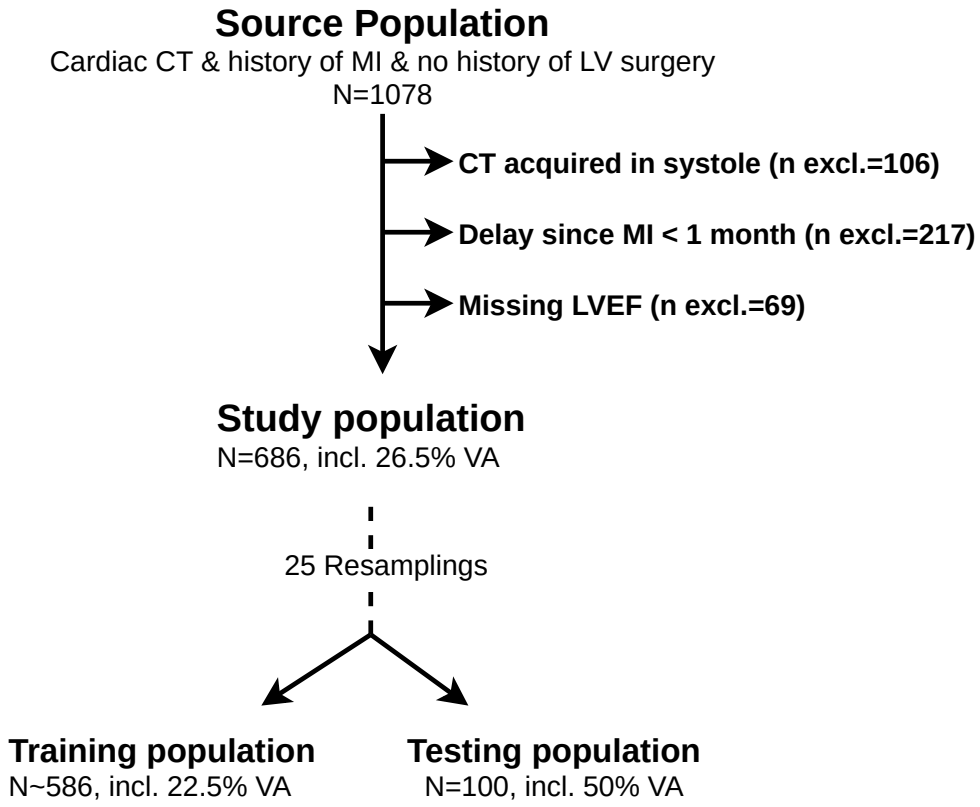


Fig. 4.8.: Population selection flowchart and bootstrap resampling.

Additionally, we compared the DL model with the LV bullseye input against the performance of the classification model with LVEF. To this end, we ran the test with a threshold model and the logistic regression model based on the LVEF value. Similarly, the LVEF models were fitted using the training dataset and evaluated with the testing dataset. To make adjustments for the class imbalance in the training dataset, the VA+ inputs were duplicated via a random sampling with replacement, using the same percentage as the VA- class. Finally, to fit the LVEF threshold model, we calculated the best cut-off value, which maximised VA classification accuracy in the training set. The best cut-off value was then applied to the testing data for evaluation.

4.3.2 Result

Table 4.1 shows the comparison of the classification accuracy of the CVAE-Class and the baseline models, after 25 resamplings. The average LVEF cut-off value calculated from the 25 resamplings was 39.4%.

The CVAE-Class model scored the highest accuracy, with an average of 75% with 4% standard deviation and the bottom end of the confidence interval higher than the mean accuracy of the baseline models. The image classifier model achieved a mean accuracy

Tab. 4.1.: The CVAE-Class and the baseline models classification accuracy.

Model	**Mean Accuracy**	**St.Deviation**	90% **CI**
CVAE-Class	**0.75**	**0.04**	**(0.74, 0.77)**
Img-Class	0.73	0.04	(0.72, 0.75)
$\text{LVEF}_{\text{LogReg}}$	0.71	0.05	(0.70, 0.73)
$\text{LVEF}_{\text{cut-off}}$	0.70	0.04	(0.69, 0.72)

CI: Confidence Interval; St.Deviation: Standard Deviation.

of 73%. Lastly, the $\text{LVEF}_{\text{LogReg}}$ model and the $\text{LVEF}_{\text{cut-off}}$ model both generated poorer results, with mean accuracy values of 71% and 70%, respectively.

The superior performance of the CVAE-Class compared to straightforward image classifier model suggests that the latent variables generated by the CVAE-Class network were relevant in improving the classification accuracy. Both models also showed a superior performance compared to the LVEF models, which further highlights the potential of the LV thickness bullseye input in identifying patients with history of VA.

The standard deviations were considerably high, ranging between 4 and 5%. Nonetheless, the stability of the standard deviations and the confidence intervals between the models rather suggest a large variability for the VA prediction between the input data, which unfortunately is expected in the clinical data.

4.4 Evaluation with Statistical Analysis

4.4.1 Experimental Setup

In this section, we aim to validate the CVAE-Class model against clinical markers (i.e., indices) by applying a traditional statistical evaluation method. First, the population (N=686) that was used in the previous section was further filtered to exclude all the patients lacking a complete set of clinical variables. Then, the study population was randomly split into training and testing population. A statistical test was used to prevent any variable distribution difference between the two populations. The train population was then used to train the CVAE-Class model, as well as to find the optimal cut-off value of the clinical variables. Finally, the models were evaluated on the test population using univariate and multivariate analysis, as well as the VA prediction accuracy. A diagram of the study methodology is shown in figure 4.9.

4.4.1.1 Statistical Analysis

The Shapiro-Wilk test was used to study the distribution normality of continuous variables. Continuous variables were expressed as mean standard deviation when normally distributed, and as median [interquartile range] ($[Q1 - Q3]$) otherwise. The relationship between continuous variables was assessed by Pearson correlation coefficients. Categorical variables were expressed in fraction ($\%$). Independent continuous variables were compared using parametric tests (i.e., unpaired Student t-test) or non-parametric tests (i.e., Mann-Whitney U test), depending on normality. Independent categorical variables were compared using Chi-Square (χ^2) tests.

The total population was partitioned into a training dataset and a testing dataset, respectively, using random sampling with a ratio of $4:1$. To ensure a consistent VA prevalence between the training and testing populations, the sampling was carried out separately in populations with and without VA. The absence of selection bias was verified by comparing clinical characteristics between the training and testing populations, including separate comparisons between the sub-populations with and without VA.

In the training population, univariable analyses were performed to identify the clinical and imaging variables associated with VA. To identify the optimum cut-off point of the significant continuous variable we applied the index of union method proposed by Unal [Unal, 2017], which bases the associated calculation on the receiver operator curve (ROC) analysis of the prediction.

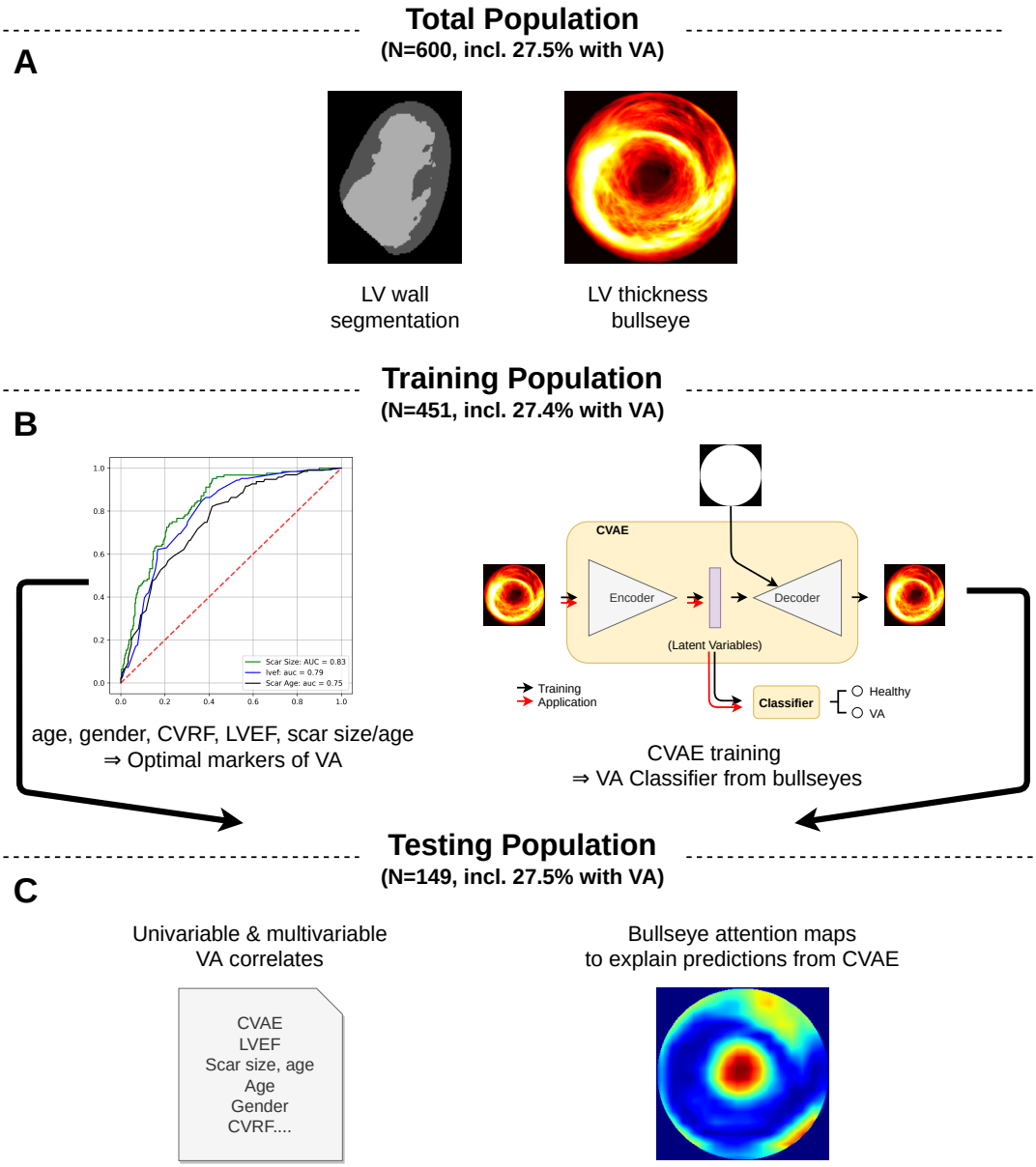


Fig. 4.9.: Study Methodology. A: CT images were segmented automatically using a trained Dual U-Net model. Automated tools were applied on the LV wall mask in order to: reformat the images into short-axis view, compute the WT, and project thickness values on a standardised bullseye map. The total population was split into training and testing populations.
B: On the training population, univariable analysis identified the clinical characteristics associated with VA, and ROC analysis the optimal cut-off values. In parallel, a CVAE model was trained to reconstruct the wall thickness WT map, then coupled to a classifier model trained to predict VA using the latent representation encoded by the CVAE.
C: The identified VA markers were applied on the testing population to study the prediction generalisability. The CVAE generated a VA prediction and an attention map highlighted the regions used for the prediction.
CVAE: conditional variational auto-encoder; CVRF: cardiovascular risk factors; LVEF: left ventricular ejection fraction; VA: ventricular arrhythmia.

Index of Union. The index of union method defines the best cut-off values as the point that would provide the highest sensitivity and specificity at the same time. The method starts with the area under the curve (AUC) value. From there, the optimal value is defined as the single point of the ROC coordinate, whose sensitivity and specificity are the closest to the AUC. Following this criterion, the optimal value c_{optim} can be calculated using the eq. 4.9, where IU denotes the index of union, AUC denotes the overall area under the curve, $SEN(c)$ and $SPE(c)$ denotes the sensitivity and specificity at value c, respectively.

$$c_{optim} = \min_{c}(IU(c)) = \min_{c}(|AUC - SEN(c)| + |AUC - SPE(c)|) \tag{4.9}$$

On the training population, the accuracy, sensibility, and specificity were assessed for all VA correlates (using the optimal cut-off value, in the case of the continuous variable), as well as for the CVAE-Class prediction.

Finally, the generalisation of the prognostic value of these markers and their cut-off values were then verified by conducting univariable analyses on the testing population. In the testing population, multivariate logistic regression analysis was performed to identify independent correlates of VA. All analyses were done using SPSS software, and by considering a $p_{value} < 0.05$ as statistically significant.

4.4.2 Result

4.4.2.1 Population characteristics

From a total of 1078 cardiac CT scans performed in patients with history of MI between 2010 and 2020, 600 subjects fulfilled the selection criteria and were used as our study population. The patient selection flowchart is shown in figure 4.10.

The indications for CT were variable: etiological diagnosis or pre-ablation assessment in patients with sustained or non-sustained ventricular arrhythmia (25%), atrial fibrillation (15%), aortic valve disease (15%), aortic disease (13%), diagnosis of coronary artery disease on native arteries or bypass grafts (11%), diagnosis of intra-LV thrombosis (8%), infectious endocarditis (4%) and other cases (9%). Population characteristics are shown in Tab. 4.2. The mean delay between the MI and the CT scan was 135 [19-216] months, and 165 (27.5%) presented with VA.

The study population was randomly divided into training and testing populations, respectively. The characteristics of the two populations are compared in Tab. 4.3. There

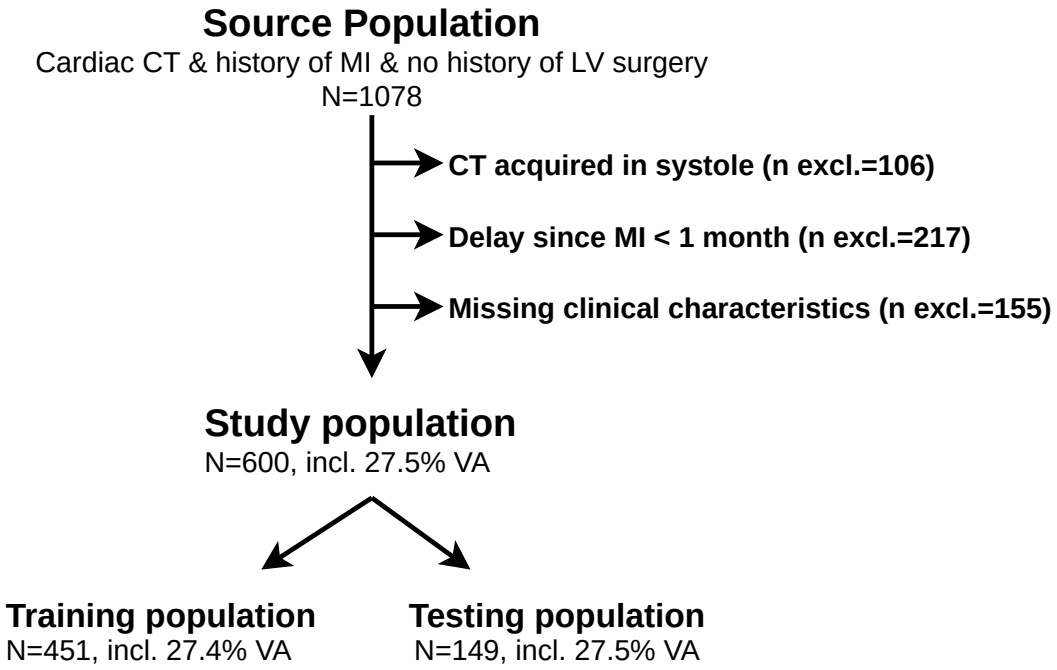


Fig. 4.10.: Population selection flowchart.

Tab. 4.2.: Population characteristics (N=600).

Variable	
Age (years)	73 [65, 82]
Female Gender	100 (17%)
Hypertension	394 (66%)
Diabetes	159 (27%)
Dyslipidaemia	436 (73%)
Smoking	270 (45%)
LVEF (%)	45 [33, 55]
Scar Size (cm^2)	56 [18, 82]
Scar Age (months)	135 [19, 216]
Unknown Scar Age	189 (32%)
Ventricular Arrhythmia	165 (27.5%)
Ventricular Tachycardia	158 (25%)
Ventricular Fibrillation	26 (4%)
Aborted cardiac arrest	26 (4%)

LVEF: left ventricular ejection fraction.

was no significant difference between the two populations, including when comparing separately subjects with and without VA.

Tab. 4.3.: Comparison between training and testing populations.

	Training pop. (N=451)	**Testing pop.** (N=149)	p_{value} (VA-)	p_{value} (VA+)	p_{value} (Total)
Age (years)	72.89 [65, 83]	72.12 [65, 80]	0.708	0.463	0.502
Female Gender	74 (16%)	26 (17%)	0.858	0.699	0.768
Hypertension	297 (66%)	97 (65%)	0.677	0.342	0.867
Diabetes	119 (26%)	97 (65%)	0.910	0.980	0.912
Dyslipidaemia	332 (74%)	104 (70%)	0.461	0.601	0.377
Smoking	206 (46%)	64 (43%)	0.626	0.303	0.563
LVEF (%)	45.25 [34, 55]	44.54 [33, 55]	0.461	0.925	0.576
Scar Size (cm^2)	56 [19, 83]	58 [19, 89]	0.433	0.839	0.655
Scar Age (months)	134 [19, 216]	140 [21, 222]	0.697	0.515	0.626
Unknown Scar Age	146 (32%)	43 (29%)	0.323	0.897	0.424
Ventricular Arrhythmia	124 (27.4%)	41 (27.5%)	N/A	N/A	N/A

Pop.: population; LVEF: left ventricular ejection fraction; N/A: not available.

4.4.2.2 Learning of ventricular arrhythmia correlates on the training population

The univariable analyses on the training population are shown in the Tab. 4.4. Among clinical characteristics, VA related with male gender ($p < 0.0001$), unknown infarct age ($p = 0.012$), LVEF ($p < 0.0001$), scar size ($p < 0.0001$) and scar age ($p < 0.0001$). On ROC analyses, LVEF, scar size and scar age showed areas under ROC curves (AUCs) of 0.79, 0.83 and 0.75, respectively, as shown in figure 4.11. The optimal cut-off value of 43% in LVEF predicted VA with an accuracy of 0.71, a sensitivity of 0.73 and a specificity of 0.70. The optimal cut-off value of $65cm^2$ in scar size predicted VA with an accuracy of 0.77, a sensitivity of 0.74 and a specificity of 0.78. The optimal cut-off value of 132 months in scar age predicted VA with an accuracy of 0.67, a sensitivity of 0.72 and a specificity of 0.65. The loss optimisation during the training of the CVAE and classifier networks is illustrated in figure 4.12. At the end of the training, the CVAE-Classifier network predicted VA with 0.87 accuracy, 0.90 sensitivity and 0.86 specificity.

4.4.2.3 Generalisation on the testing population

Univariable analyses performed on the testing population by applying the optimal cut-off values learned on the training population are shown in Table 4.5. As in the training population, VA related to male gender ($p = 0.045$), LVEF ($p < 0.0001$), scar size ($p < 0.0001$) and scar age ($p < 0.0001$). CVAE predictions were also significantly related to VA ($p < 0.0001$), confirming the model's generalisability.

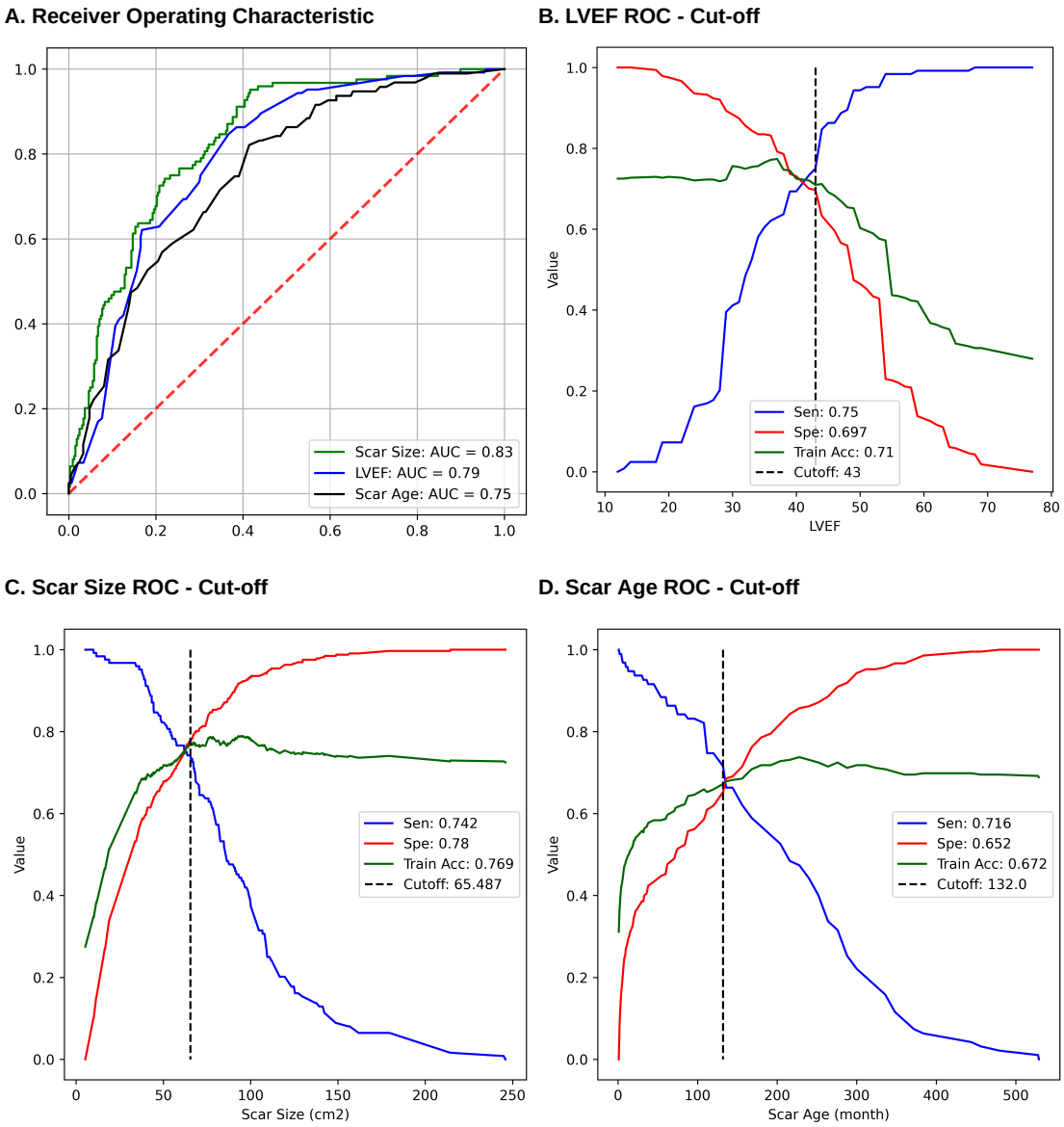


Fig. 4.11.: ROC analysis for the prediction of ventricular arrhythmia using LVEF and scar size. A: ROC curves. B: Definition of the optimal cut-off value for LVEF. C: Definition of the optimal cut-off value for scar size. D: Definition of the optimal cut-off value for scar age.

Tab. 4.4.: Univariable correlates of ventricular arrhythmia in the training population.

	Training pop. (N=451)	**VA-** (N=327)	**VA+** (N=124)	p_{value}
Age (years)	73 [65, 83]	74 [65, 84]	71 [64, 77]	0.498
Female Gender	74 (16%)	67 (20%)	7 (6%)	< 0.0001
Hypertension	297 (66%)	220 (67%)	77 (62%)	0.300
Diabetes	119 (26%)	89 (27%)	30 (24%)	0.515
Dyslipidaemia	332 (74%)	242 (74%)	90 (73%)	0.759
Smoking	206 (46%)	142 (43%)	64 (52%)	0.119
LVEF (%)	45 [34, 55]	49 [40, 55]	36 [30, 45]	< 0.0001
Scar Size (cm^2)	55 [19, 83]	42 [15, 63]	91 [62, 110]	< 0.0001
Scar Age (months)	133 [19, 216]	101 [8, 156]	206 [108, 288]	< 0.0001
Unknown Scar Age	146 (32%)	117 (36%)	29 (23%)	0.012

VA+: patient with ventricular arrhythmia; VA-: patient without ventricular arrhythmia; LVEF: left ventricular ejection fraction.

A multivariable model was constructed using these 4 variables included in Tab. 4.6. Results showed that the parameters most strongly associated with VA were the CVAE prediction and scar age, whose collinearity rendered non-significant the contribution of LVEF, scar size and male gender.

The diagnostic performance of CVAE prediction, LVEF and scar size, as assessed on the testing population are shown in Tab. 4.7. The CVAE predicted VA with 0.79 accuracy, 0.73 sensitivity, and 0.82 specificity. Once the model was trained, the computing time of model prediction including the generation of the class activation mapping is less than 1s

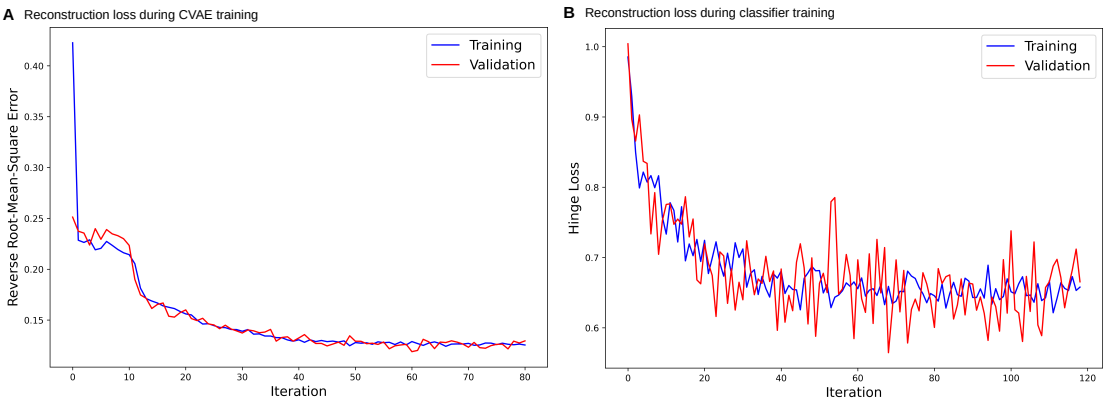


Fig. 4.12.: Loss optimisation during the training of the CVAE and classifier networks. A: Optimisation of the reconstruction loss when training the CVAE; B: Optimisation of the prediction loss when training the classifier.

Tab. 4.5.: Univariable correlates of ventricular arrhythmia in the testing population.

	VA- (N=108)	**VA+** (N=41)	p_{value}
Age	73 [66, 81]	70[61, 78]	0.106
Female Gender	23 (21%)	3 (7%)	0.045
Hypertension	75 (69%)	22 (54%)	0.071
Diabetes	30 (28%)	10 (24%)	0.677
Dyslipidaemia	76 (70%)	28 (68%)	0.805
Smoking	44 (41%)	20 (49%)	0.376
Unknown Scar Age	33 (31%)	10 (24%)	0.458
Scar Age > 132 months	24 (36%)	22 (84%)	< 0.0001
LVEF < 43%	37 (34%)	29 (71%)	< 0.0001
Scar Size > $65.49cm^2$	37 (34%)	29 (71%)	< 0.0001
CVAE-Class	20 (19%)	30 (73%)	< 0.0001

VA+: patient with ventricular arrhythmia; VA-: patient without ventricular arrhythmia; LVEF: left ventricular ejection fraction; CVAE-Class: conditional variational autoencoder classifier model.

Tab. 4.6.: Multivariable correlates of ventricular arrhythmia in the testing population.

	Standardised β	**Partial** R^2	p_{value}
CVAE-Class	0.409	0.291	< 0.0001
Scar Age > 132 months	0.262	0.061	0.004
Scar Size > 65.49 cm^2	0.175	0.013	0.080
LVEF < 43%	-0.125	0.008	0.239
Male Gender	-0.063	0.004	0.446
	Model R^2	**0.375**	

CVAE-Class: conditional variational classifier model; LVEF: left ventricular ejection fraction.

(once the model is loaded). Combining with the image processing step, the completion of all steps in the pipeline for each testing case was less than 3 minutes.

4.5 Model Explainability Result

To explain CVAE-Class predictions, the attention maps reconstructed using the Grad-CAM++ method were qualitatively analysed. Examples of attention maps in patients

Tab. 4.7.: Diagnostic performance of imaging markers of ventricular arrhythmia in the test population. The rows are sorted in the descending order of the accuracy.

	Accuracy	Sensitivity	Specificity
CVAE-Class	0.792	0.732	0.815
Scar Size $>$ 65.49 cm^2	0.725	0.683	0.741
Scar Age $>$ 132 months	0.698	0.839	0.640
LVEF $<$ 43%	0.671	0.707	0.657

CVAE-Class: conditional variational autoencoder classifier model; LVEF: left ventricular ejection fraction.

with and without VA are shown in figure 4.13 and in figure 4.14. Figure 4.13 shows the example correction predictions of the CVAE-Class model, where the model is shown to be robust. The true positive case in figure 4.13.A.2. shows that the model is sensitive to the thinning location even when the scar has a small size. The model is also shown to be robust against the extended scar regions, which are closer to the apex, in the true negative cases, as seen in figure 4.13.

However, the model was trained based on the LV thickness; thus, it was prone to the false prediction cases related to the thickness value, as shown in figure 4.14. The false positive cases also highlight the limit of the GradCAM++ method, as the thinning regions are stills activated. In contrast, in other image classification contexts where the activated regions may be used to manually reject the prediction, the manual validation is not as trivial in the case of VA predictions.

The GradCAM++ method is designed for the case-by-case visual explanation of the model prediction, and thus it does not allow quantitative analysis of the model general tendency. For example, additional analysis would be needed to determine whether the thinning in the first quadrant of the bullseye weights more in the VA+ class than the third quadrant. Nevertheless, the visual explanation did confirm that the most influential regions to predict VA co-localised with the scar region. It also suggested that the network operated as a filter to distinguish scar from physiologically thin regions (e.g. those located at the LV base next to the mitral valve). This may explain why the model outperformed scar size measured using a thickness-based threshold. Nonetheless, the high gradient regions were still considerably large, and could not be used to pinpoint discrete arrhythmogenic patterns within the scar.

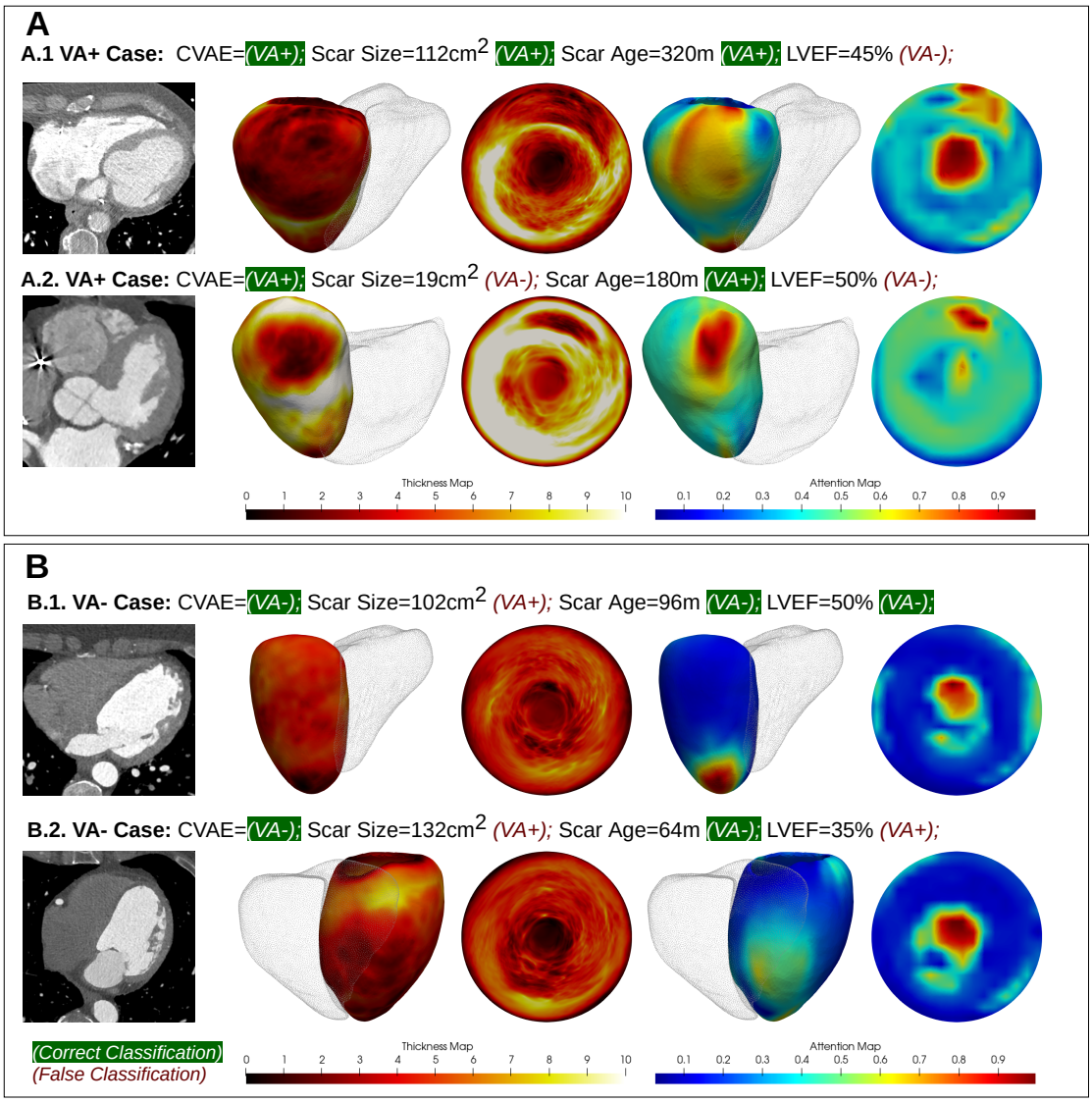


Fig. 4.13.: Explanation of model's predictions by attention map. A: 2 patients, correctly classified as VA- positive by the CVAE model. B: 2 patients, correctly classified as VA-negative by the CVAE model. Left column: original CT image. Mid column: 3D meshes and bullseye maps of LV WT. Right column: 3D

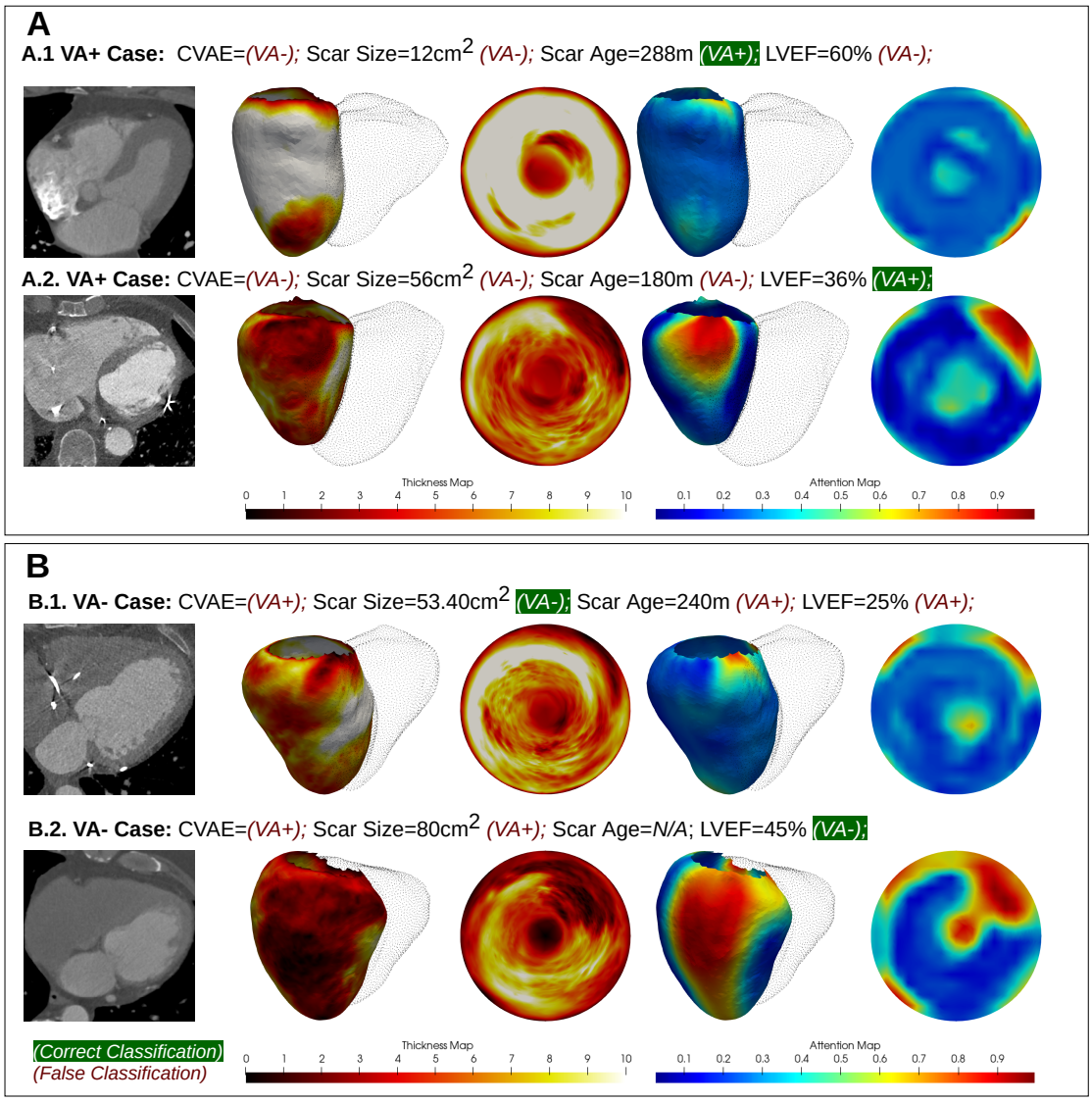


Fig. 4.14.: Explanation of model's predictions by attention map. A: 2 patients, falsely classified as VA- positive by the CVAE model. B: 2 patients, falsely classified as VA-negative by the CVAE model. Left column: original CT image. Mid column: 3D meshes and bullseye maps of LV WT. Right column: 3D

4.6 Discussion

To our knowledge, this work is the first to study the relationship between CT-defined WT heterogeneity of the LV in the presence of scar-related VA. To sum up, we introduced a fully automated method for the extraction, quantification and mapping of LV WT from CT images, and demonstrate that WT heterogeneity can be analysed by deep learning models to generate VA predictions. Analysing a large retrospective database of 600 patients, the proposed approach outperformed the accuracy of existing risk markers for the identification of patients with VA. These results outline the strong potential of cardiac CT to extract scar features applicable to risk stratification, confirming the role of wall thinning in the development of arrhythmogenic substrates after MI.

4.6.1 Methodological approach

In the present study, we chose to use a CVAE to reduce the dimension of the problem in a latent representation. This is justified by a prior work from our group using the same training database, in which we demonstrated that a classifier model fed with such latent variables provides a more accurate prediction compared to methods classifying VA/non-VA patients directly from original bullseye maps [Ly, 2021]. Note that a similar approach was previously applied to the diagnosis of hypertrophic cardiomyopathy [Oktay, 2018].

Another advantage of the CVAE is its generative aspect. This aspect was not specifically studied in the current work; however, the latent space could be explored to reconstruct archetypal images of a given condition in order to generate pathophysiological hypotheses, as proposed by Biffi et al. for hypertrophic cardiomyopathy [Biffi, 2020].

Last, the generation of attention maps to explain each prediction shows several benefits. Firstly, it can be used to verify that the regions used for the prediction actually matched well the infarct location, thus increasing the confidence and transparency of the prediction result. Similar approaches were proposed in other domains of medical imaging, including in the detection of pulmonary nodules [Joshua, 2021] or in the diagnostic of osteoarthritis [Karim, 2021]. Secondly, these maps can be used to explain the model performance and potentially to identify arrhythmogenic patterns within the scar area, which could be used as preferential targets for guiding the ablation procedure.

4.6.2 Ventricular Arrhythmia prediction

Our results confirm the potential of scar characterisation for the prediction of post-MI VA. On our testing population, the proposed model outperformed LVEF (i.e, the current criterion used for primary prevention). These results are consistent with several prior

studies that have used CMR methods based on late gadolinium enhancement to assess the scar size or heterogeneity [Kwon, 2009; Heidary, 2010; Zegard, 2021]. However, these approaches required human intervention for the scar segmentation, while the reproducibility of the proposed quantification is still debated [McAlindon, 2015].

The method introduced in the present study is fully automated, and based on an imaging modality (i.e., CT) known for being widely accessible and reproducible. Our results showed that model predictions were strongly associated with VA, and that the performance was robust to generalisation. Moreover, multivariate analysis showed that the model predictions remained independently related to VA, while LVEF was not. This suggests that WT contains an implicit information on LV systolic dysfunction.

Notably, the predictions from the model outperformed a simple thickness thresholding, confirming the role of thickness heterogeneity within the scar region in the formation of an arrhythmogenic substrate. This further suggests a potential adaptability of the model to the physiological LV thickness heterogeneities such as thinning at LV base and apex, thicker wall on the septum than on the free wall.

It should also be noted that the optimal LVEF threshold identified in our population is by 8% higher than the current recommendations for the selection of ICD candidates (43% vs 35%). This could be explained by the retrospective nature of our dataset, and by a potential selection bias related to the specific indications to perform cardiac CT in patients with ischaemic heart disease in our centre. This could be explained by the fact that the proposed cut-off value resulted from an optimisation process, while the recommended cut-off of 35% was adopted because it was part of the inclusion criteria of large clinical trials that validated ICD therapy [Buxton, 1999], with no similar optimisation.

Lastly, the different types of arrhythmias integrated in our composite criterion (e.g., VT, VF, and aborted sudden death), although all potentially lethal, do not correspond to similar mechanisms and could occur within structurally distinct substrates. In our dataset, sustained VT was by far the most frequent mechanism, which is consistent with prior series in patients with ischaemic cardiomyopathy [Brugada, 2001]. Therefore, the model was essentially trained to detect the substrate associated with scar-related re-entrant VT. Using the same latent space to identify distinctive features of the VF substrate would be extremely pertinent, since the generative nature of the model could be used to produce as well as to compare archetypical WT distributions of the two mechanisms. To this end, however, a database comprising a larger number of VF episodes would be required.

4.6.3 Explanation of model performance

To avoid the "black box" effect often associated with artificial intelligence tools in medicine, we opted for generating attention maps explaining the origin of each prediction in the image. As explained above, these maps were useful to increase clinical confidence in the prediction by demonstrating that the VA risk was indeed identified in the scar region. They also provided explanation as to why the model performed better than a simple thickness thresholding.

Indeed, LV wall thinning was shown to be related to post-infarction scar [9], but scar quantification based on thickness measurements may be hampered by the physiological heterogeneity of LV thickness. The basal and apical regions of the LV are physiologically thin, and the WT is also heterogeneous around the LV circumference (i.e., thicker on the septum or at the insertion of papillary muscles).

Our results indicate that as opposed to thickness thresholding, the proposed model integrates this physiological thickness heterogeneity and may provide more robust scar characterisation. Nevertheless, the regions of high gradient on the attention maps were broadly covering the scar region, and did not allow for detecting discrete scar sites particularly prone to arrhythmogenesis. This was certainly caused by the loss in spatial resolution when encoding into the latent space.

Prior studies comparing WT maps with electrophysiological mapping during post-MI VT showed that the critical isthmus of the VT circuit often colocalises with channels of relatively preserved thickness protruding into or even traversing severely thinned scar. However, the size of these channels (with average width of 9mm and surface < 1.5cm2) [Takigawa, 2019] might be inaccessible to the current resolution of our model. Thus, in its current state, the model can be successfully applied to predict the presence of VA or to localise scar regions, but could not be used to provide precise targets for catheter ablation.

4.6.4 Limitations

4.6.4.1 CVAE-Class Model

Due to the compression of the CVAE, the reconstructed bullseye map did not contain some high frequency details, as shown in figure 4.15.A. While this could be due to the failure of the decoder branch, it could also suggest that the latent vectors did not hold necessary information to decode the missing details. Overall, the current framework focused on the classification and did not directly use the reconstructed output, ensuring that the latent vector contains heterogeneity information could lead to better classification accuracy.

We analysed the latent variables of the test data using the partial least square regression, as seen in figure 4.15.B, in which we can observe a partial overlap between the two classes. Adding constraints to increase the separation of the latent variables could allow for a more robust classification model. Moreover, in a well separated latent space, we could exploit the latent variables to generate the LV bullseye map of the different classes. This would allow us to display the personalised features that influence the classification.

Since our current model was only trained with the LV thickness map, its accuracy was ultimately tied to the cases where the VA was related to the wall thinning, as illustrated by the false predictions in figure 4.14. The subtle differences between the bullseye could be the key to a more accurate classification, therefore it is crucial to improve the latent representation. On the other hand, integrating the clinical information into the current framework can be the direct answer to the single-input limitation.

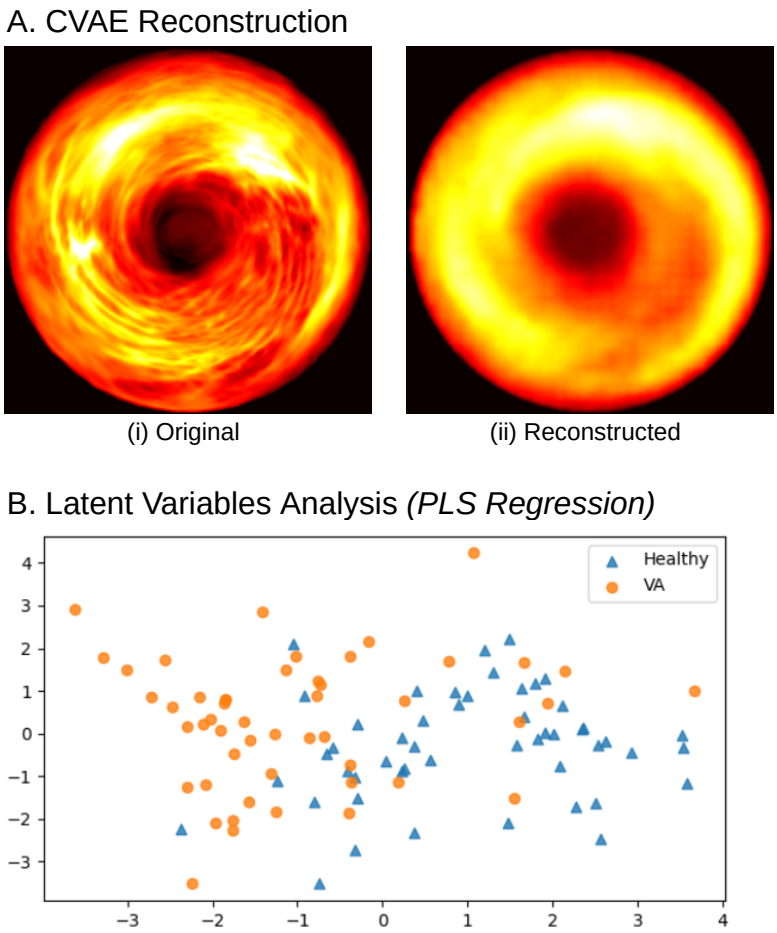


Fig. 4.15.: A. Image reconstruction; B. Latent variables analysis.

4.6.4.2 Dataset

This is a retrospective study describing the association between imaging features and past arrhythmia events. Therefore, the prognostic value of the proposed risk markers remains to be demonstrated in a prospective fashion. The absence of arrhythmia could not be systematically verified using long term Holter monitoring, and it is possible that some patients in the no VA group might have presented undetected episodes. However, this is quite unlikely, given the criteria used to define VA episodes. This retrospective work did not allow for an exhaustive collection of clinical characteristics. In order to avoid limiting the inclusion criteria, only the most important characteristics were collected, while additional prognostic markers or confounding factors may have been omitted.

Finally, we acknowledge that the radiation exposure associated with the CT modality may limit future clinical translation, particularly for risk stratification purposes. However, clinical indications for cardiac CT in asymptomatic patients are already present (i.e. coronary calcium score), and in addition a new generation of CT systems exposing to less radiation is currently being introduced in the clinical field [Leng, 2019].

4.7 Conclusion

In this chapter we introduced a fully automated method for the analysis of left ventricular WT defined from CT images, leveraging on the framework of generative models to obtain VA predictions in patients with history of MI. The model was trained, validated and tested on a large retrospective database comprising 600 patients. Our results showed that WT heterogeneity outperforms the accuracy of existing risk markers for the identification of patients with VA. This outlines the potential of cardiac CT for the extraction of robust SCD risk stratifiers and confirms the role of WT in post-MI arrhythmogenesis.

5 Ventricular Arrhythmia Prediction using Interpretable Graph Convolutional Network

Contents

In this chapter we proposed a novel prediction method that can improve the VA classification accuracy obtained with the CVAE-Class model described in the previous chapter 4. While our broad goal is to still to base the classification on the LV thickness map, here we particularly aim to address the limitations of the previous prediction pipeline, such as those the issues related to the grid-based calculation of the convolutional filter of the CNN and those due to the flattening distortion of the LV bullseye input.

Specifically, in this study we developed an automatic pipeline for VA prediction based on CT images, using a Graph Convolutional Network (GCN). The prediction is similarly based on the theory that the scar heterogeneity observable as a myocardium thinning in CT images, plays an important role as an anatomical substrate in VA mechanism.

The pipeline includes the same steps as described in chapter 3: the segmentation of LV masks from the input CT image, the short-axis orientation reformatting, LVMYO thickness computation and mid-wall surface mesh generation. An average LV mesh was computed and fitted to every patient in order to use the same number of vertices with point-to-point correspondence. The GCN model was then trained using the thickness value as the node feature and the atlas edges as the adjacency matrix. This allows the model to process the data on the 3D patient anatomy and bypass the “grid" structure limitation of the traditional convolutional neural network.

The model was trained and evaluated on a dataset of 600 patients (27% VA), using 451 (3/4) patients for training and 149 (1/4) patients for testing data, respectively. The evaluation results showed that the graph model (81% accuracy) outperformed the clinical baseline (67%), the left ventricular ejection fraction, and the scar size (73%).

We further studied the interpretability of the trained model using LIME (local interpretable model agnostic explanation) and integrated gradients and found promising results on the personalised discovering of the specific regions within the infarct area related to the arrhythmogenesis. With this respect, figure 5.1 illustrates the summary of the study investigated here.

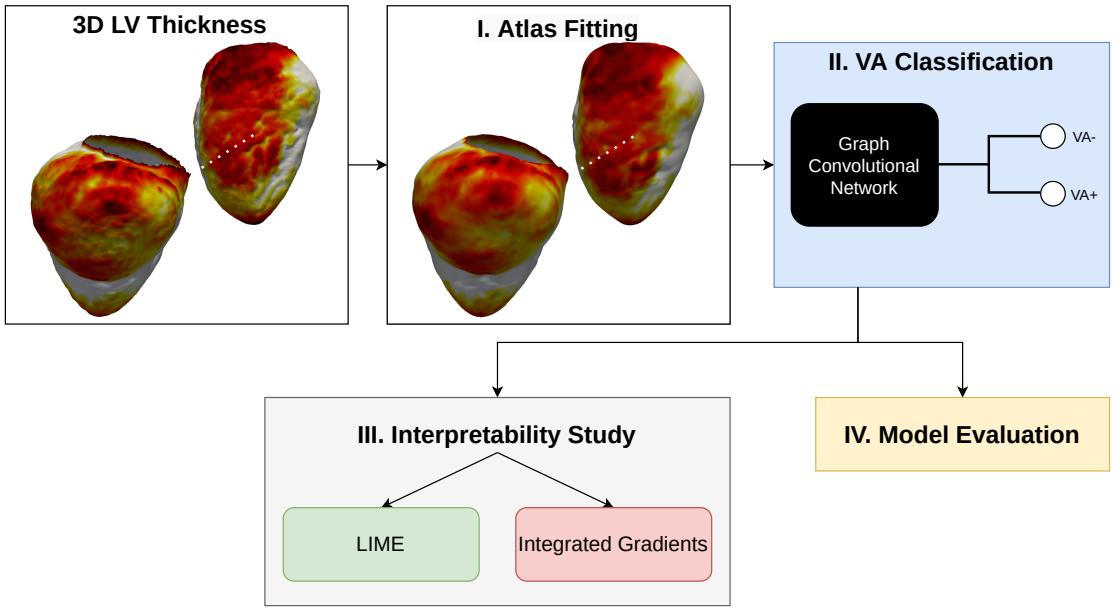


Fig. 5.1.: The chapter road map to build an interpretable VA graph classification model. The LV atlas fitting is detailed in section 5.2.1, and The graph convolutional network architecture is described in section 5.2.2. The model is evaluated in section 5.3 and 5.4. We study the model interpretability in section 5.2.3 and 5.4.3.

Our main contribution in this chapter are:

- We proposed a novel VA prediction pipeline from CT images, which achieved better accuracy than the CVAE-Class model.

- We showed that the point-to-point correspondence of the LV midwall mesh, through atlas registration, is crucial to the fitting of the GCN model and allow to produce more accurate prediction.

- We proved that the index pooling method allows the GCN model to better integrate spatial information and to achieve superior prediction accuracy.

- We proved that fitting the DL model directly using the surface mesh of the LV using the GCN model resulted in a more robust VA classification model than using the conventional CNN models.

- We proposed an adaptation of the LIME method on the graph network.

- The interpretability studies in this study using LIME and integrated gradients on the GCN model allow for a more discrete localisation of the activation regions, which is a step closer to the identification of the arrhythmogenesis site.

5.1 Introduction

In recent years, Deep Learning (DL) models such as the convolution neural network (CNN), has made a remarkable impact in medical image processing, ranging from semantic segmentation to diagnosis and outcome prediction. Notably, the grid-like property of the CNN can limit its efficacy on the organs with specific 3D anatomy, such as the LV myocardium. In the previous chapter, we proposed a transformation of the CT input image into a 2D bullseye representation of the LV thickness, for the purpose of removing the endocardial blank voxels. However, the flattening of the 3D surface mesh led to a distortion of the 3D thickness map heterogeneity, impacting the subsequent fitting of the DL model. On the other hand, graph neural networks have gained more applications in medical image processing and bioinformatics [Zhang, 2021], thanks to their adaptability to specific data geometry.

In this study, we investigated the graph-level classification task using a model built with graph convolutional network (GCN) layers [Kipf, 2017], and the LV thickness map as input. Moreover, we studied the interpretability of the graph model using LIME [Ribeiro, 2016] and integrated gradients [Sundararajan, 2017] to locate the specific regions contributing to the VA prediction and to gain better insight on the explicit arrhythmogenic regions from the model perspective.

5.2 Method

In this study, we used the LV mid-wall meshes generated in the chapter 3. To prepare the meshes as input to the graph network, we applied atlas mesh fitting to ensure a point-to-point correspondence between the inputs. Next, a graph classification network was built using the atlas mesh and the thickness value as input. The overview of the proposed VA prediction pipeline is shown in figure 5.6. Note that the image processing pipeline and the model architecture are described in section 5.2.1 and section 5.2.2, respectively.

5.2.1 LV Mid-wall Mesh Processing

As detailed in chapter 3, the LV mid-wall mesh was generated using marching cube algorithm following with the uniform re-meshing using the ACVD algorithm. Although, the input mid-wall images were SAX reoriented and isotropically resampled, which preserved the rotation and spacing consistency, there is not point-to-point correspondence between the output meshes. To this end, we used the open-source software Deformetrica [Bône, 2018], which allows for the deterministic atlas fitting of the 3D meshes based on a large deformation diffeomorphic metric mapping (LDDMM) framework. The average mesh was generated for all of the available meshes, after which the thickness value was remapped to the registered average mesh, which was next used as input to the graph model.

Figure 5.2 summarises the processing pipeline used in this project, from the initial CT image to the atlas registered LV mid-wall mesh.

5.2.1.1 Deformetrica: Deterministic Atlas

The LDDMM framework allows for the comparison between shapes, either in 2D or 3D, using diffeomorphic transformations of the ambient space between these objects. The diffeomorphism theory in the ambient space is largely influenced by the mathematician biologist D'Arcy Thomson and his works on "the Theory of Transformation, or the Comparison of Related Forms" [Thompson, 1942], where two shapes can be transformed into one another through the transformation of the space surrounding the object, namely, the ambient space. The LDDMM framework has two distinct methods of shape analysis: the parameterisation of a large family of transformations and the computation of distances between the objects.

Under the framework of Deformetrica, the transformation map $\Phi_{q,\mu}$ is parameterised by a set of the control points $(\mu_i)_{i=1,\ldots,n}$ and a corresponding set of momentum vectors

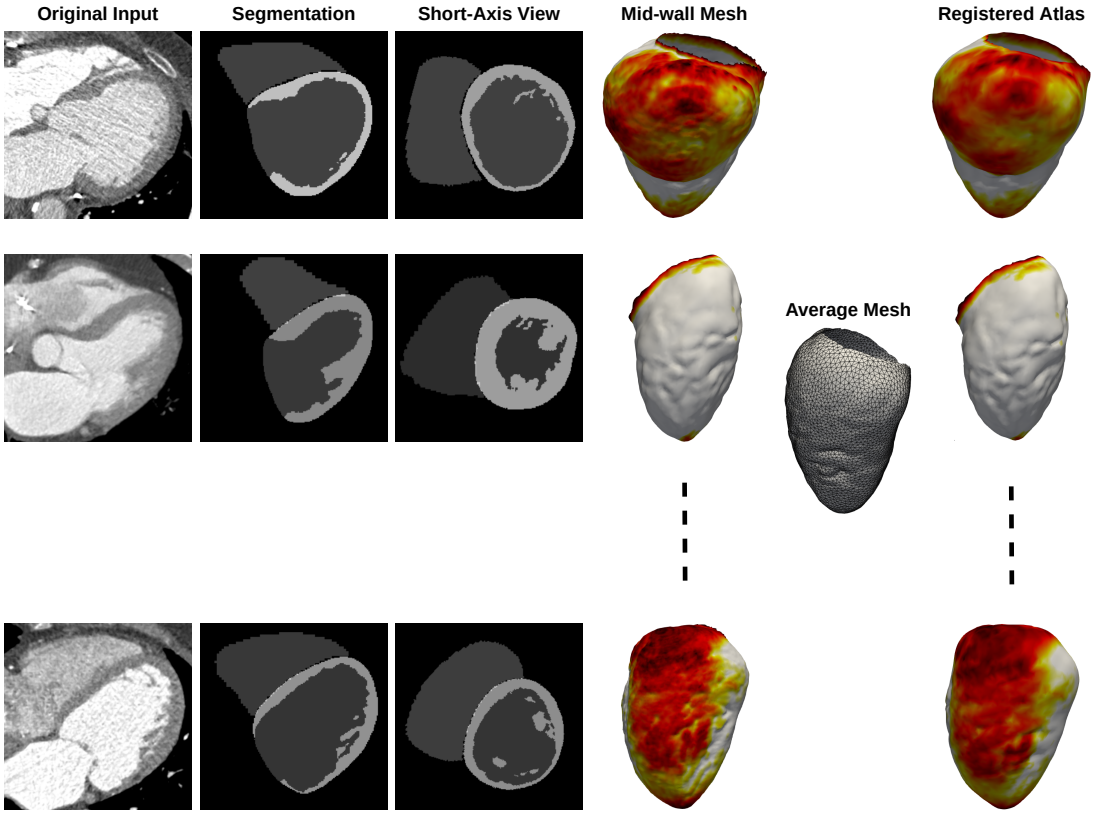


Fig. 5.2.: Processing pipeline to generate the graph network dataset from cardiac CT images. The individual LV surface meshes were generated before the average mesh was fitted by using a deterministic atlas method. The thickness values were then transferred to the registered atlas mesh via linear interpolation.

$(\mu_i)_{i=1,\dots,n}$. The parameterisation is done through the iteration of a dynamic velocity field generated using the with the following equation:

$$X(x) = \sum_{i=1}^{p} K(x, q_i) \cdot \mu_i \tag{5.1}$$

where $X(x)$ denotes the vector at point x and $K(x,q) = exp\left(-\frac{||x-q||}{\sigma^2}\right)$ denotes the Gaussian kernel of the width σ, which is used to control the finest of the registration. The smaller σ outputs a shape registration with finer precision.

To calculate the distance between the control points, the Euclidean distance (l^2) was used for the image data. For the mesh objects, Deformetrica adopted the calculation of distances using the centres of triangles $(c_k)_{k=1,\dots,r}$ and the edge normals $(n_k)_{k=1,\dots,r}$ using the Eq. 5.2.

$$d\left((n_k^\alpha, c_k^\alpha)_{p=1,\dots,r}, (n_l^\beta, c_l^\beta)\right)^2 = \sum_k \sum_l K_w(c_k^\alpha, c_l^\beta) \cdot (n_k^\alpha)^T n_l^\beta \tag{5.2}$$

The deterministic atlas model computes the average shape T and a set of transformation maps $(\Phi_i)_{i=1,\dots,n}$ for each shapes $(S_i)_{i=1,..,n}$, by minimising the following function:

$$L(T, q, \mu_{i=1,\dots,n}) = \sum_i d(\Phi_{q,\mu_i} \circ T, S_i)^2/\sigma_\epsilon^2 + R(q, (\mu_i)_{i=1,..,n}), \tag{5.3}$$

$$\text{where} \quad R(q, (\mu_i)_{i=1,\dots,n}) = \sum_i \mu_i^T K(q,q)\mu_i \tag{5.4}$$

The $R(\cdot)$ in the Eq. 5.4 represents the regularisation term, which penalises the kinetic energy of the deformation.

To prepare the LV surface meshes dataset as the input to the deterministic atlas function, we reapplied the uniform re-meshing method to control the number of the mesh vertices. The number of vertices would eventually impact the graph model architecture. Rigid registration of the meshes was done prior to the atlas fitting. As the rotation and size were already registered, we translated the centre of mass of all the meshes toward a reference centre, calculated from a randomly selected mesh.

We fitted all the available meshes (training and testing) into the deterministic atlas model. The uniform meshing was re-applied to create the dataset with different number of vertices for the graph model. At inference, the atlas mesh could be directly registered to the new mesh using affine registration. Figure 5.3 illustrates the atlas registration done by Deformetrica.

5.2.2 Graph Convolutional Network Model

We built the model using the GCN layer [Kipf, 2017], which used the following convolutional operation:

$$X' = \hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{\frac{1}{2}}XW + b\,, \tag{5.5}$$

where $X' \in \mathbb{R}^{N\times F}$ is the output of N nodes and F filters, $X \in \mathbb{R}^{N\times C}$ is the input of C features, $W \in \mathbb{R}^{C\times F}$ is the weights, b is the bias, $\hat{D} \in \mathbb{R}^{N\times N}$ the degree matrix, and $\hat{A} \in \mathbb{R}^{N\times N}$ is the self-loop adjacency matrix ($\hat{A} = I_N + A$).

Using the Eq. 5.5, the information can be aggregated between nodes, thanks to the adjacency matrix. The self-loop workaround allows the integration of the node's own information into the output X'. Starting with the direct adjacent node in the first layer

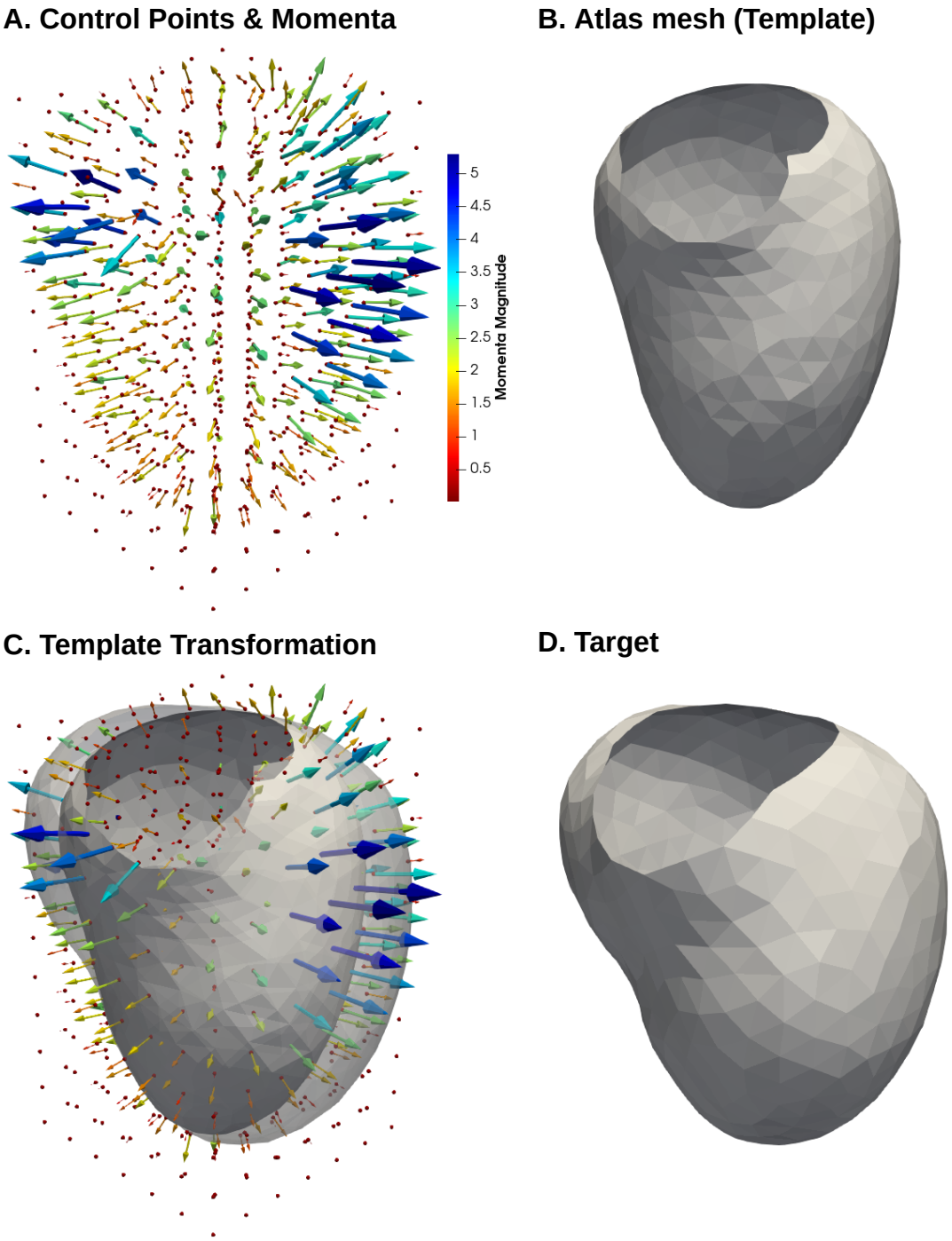


Fig. 5.3.: Deformetrica atlas registration. A. The set of control points and momentum to transform the template toward the target mesh. The arrows are scaled based on the magnitude of the momenta. B. The average/atlas mesh, also called the template mesh in Deformetrica. C. The template transformation using the vector field generated from the control points and the corresponding momenta. D. The target mesh.

of GCN, the information from the nodes of higher degree connection can be aggregated to the node through the subsequent layers of GCN. The information on aggregation for the GCN model is illustrated in Figure 5.4.

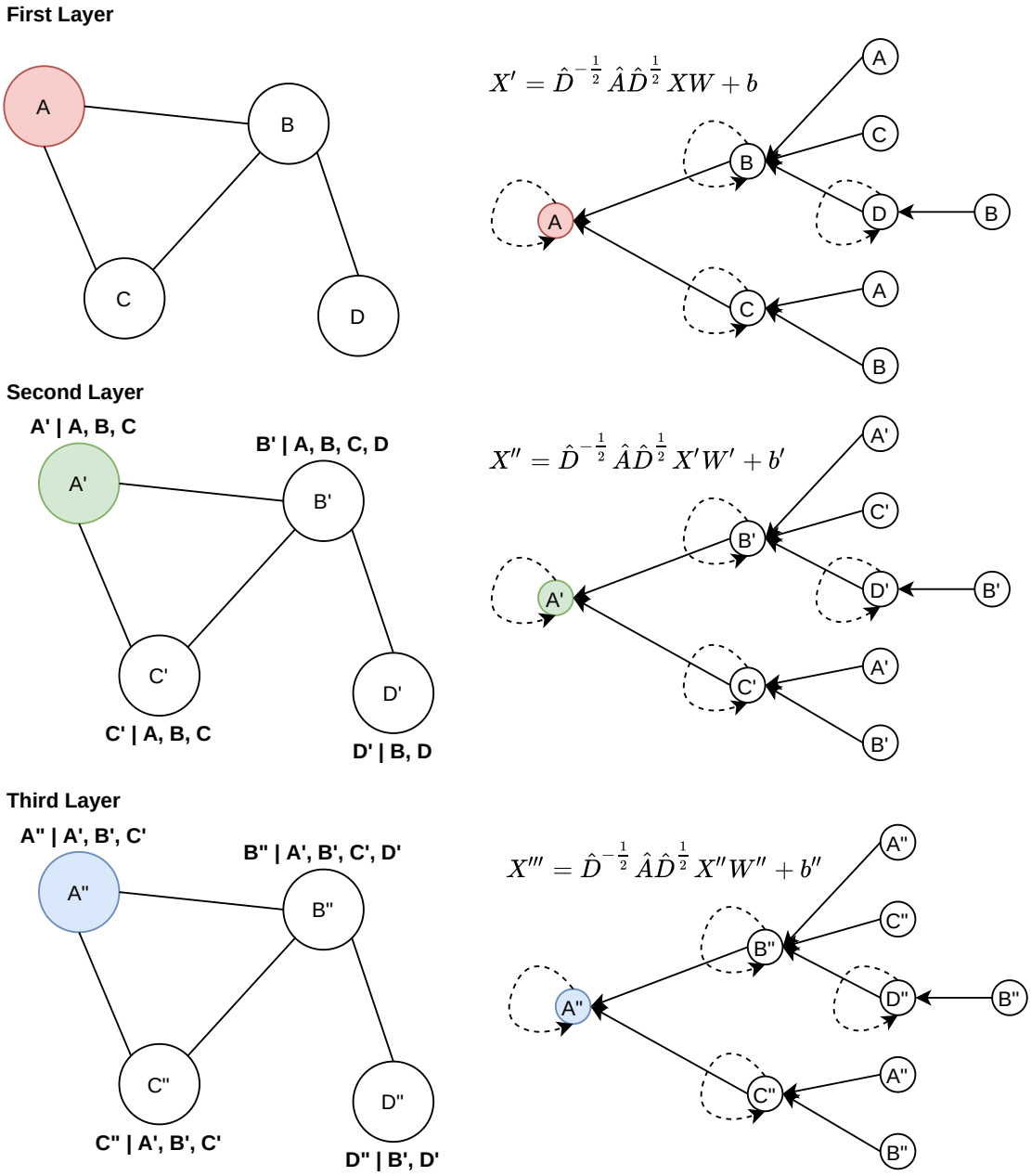


Fig. 5.4.: The information aggregation through GCN model. The adjacency matrix $\hat{A}$ defines the node relation and is constant throughout the layer. Note that the information of the node D is aggregated to the node A (A") by the third layer, through B'. Similarly, the information for the node A is also aggregated to D (D"). The dotted arrow denotes the self-loop aggregation.

We considered the LV mid-wall mesh as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes $v_i \in \mathcal{V}$ and edges $(v_i, v_j) \in \mathcal{E}$. We used the thickness value of each node v_i as the input feature $X \in \mathbb{R}^{N\times 1}$. Since the inputs were uniformly re-meshed, no edge feature was used. The adjacency matrix A was extracted from the edges of atlas mesh and used for each input. Figure 5.5 shows the calculation of the adjacency matrix from an exemplary LV surface mesh. The LV mesh could then be represented as a graph $\mathcal{G} = (X, A)$.

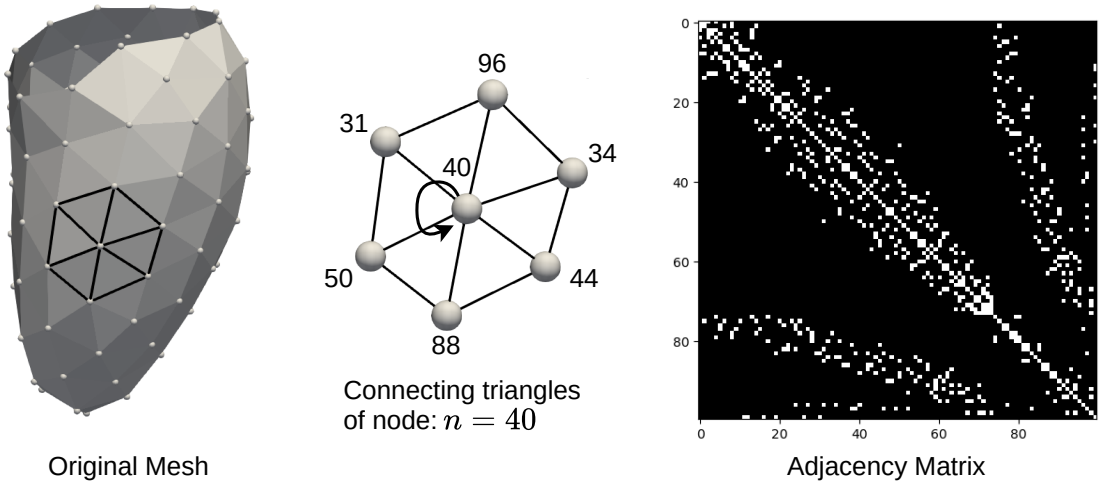


Fig. 5.5.: The extraction of the adjacency matrix from the LV surface mesh. The self-loop can be seen as the entries in the main diagonal of the adjacency matrix.

5.2.2.1 Index Pooling.

The GCN layer calculated the aggregated output based on the point-wise feature input and its adjacency matrix; thus, the downsampling methods used in the CNN (such as strided convolution or MaxPooling layers) are not applicable to the GCN architecture. Therefore, we employed the pooling method as proposed by [Simonovsky, 2017], where a coarse graph was generated in order to pre-define the corresponding pooling index and the new adjacency matrix during the model construction. In our case, the coarse graph was generated using the ACVD uniform meshing with reduced number of points, and the corresponding pooling indexes were defined using k-dimensional trees nearest neighbour search (KD trees NN) algorithm. The pooling index was defined using the following steps:

- Build the KD trees using the point coordinates of the coarse mesh.
- Iterate through all the points in the original mesh and assign the closest node as the pooling index.
- When running the model, apply the pooling function (ex. max, min or average pooling) among the output features belonging to the same pooling index.

Here we built the index pooling layer using max pooling method. The graph coarsening and pooling indexes searching are computationally expensive, which is unfit for online processing. This pooling method is more suitable with a point-to-point correspondent graph input, as the pooling indexes of the atlas mesh can be applied to all of the input graphs. The index pooling steps are shown in figure 5.6.III.

5.2.2.2 GCN-Pooling: Model Architecture.

The model architecture is shown in figure 5.6.II. Starting with an input graph of N nodes, the model was built using two consecutive GCN layers, followed by an index pooling layer, which reduced the aggregated graph to N_{dwn} nodes. The pooling output was then passed through three more GCN layers, before feeding the fully-connected network (FCN) classification block. We set the filter size to 64 for all the GCNs, except for the last GCN for which the filter was set to 1. The classification block was built with 4 fully-connected layers with $[128, 64, 32, 2]$ units, respectively. We applied a linear rectifier activation to the first 3 layers and the softmax activation for the output layer. The model was trained using the binary cross-entropy loss. The model was built using tools provided by Spektral [1] (GCN layer) and Tensorflow [2].

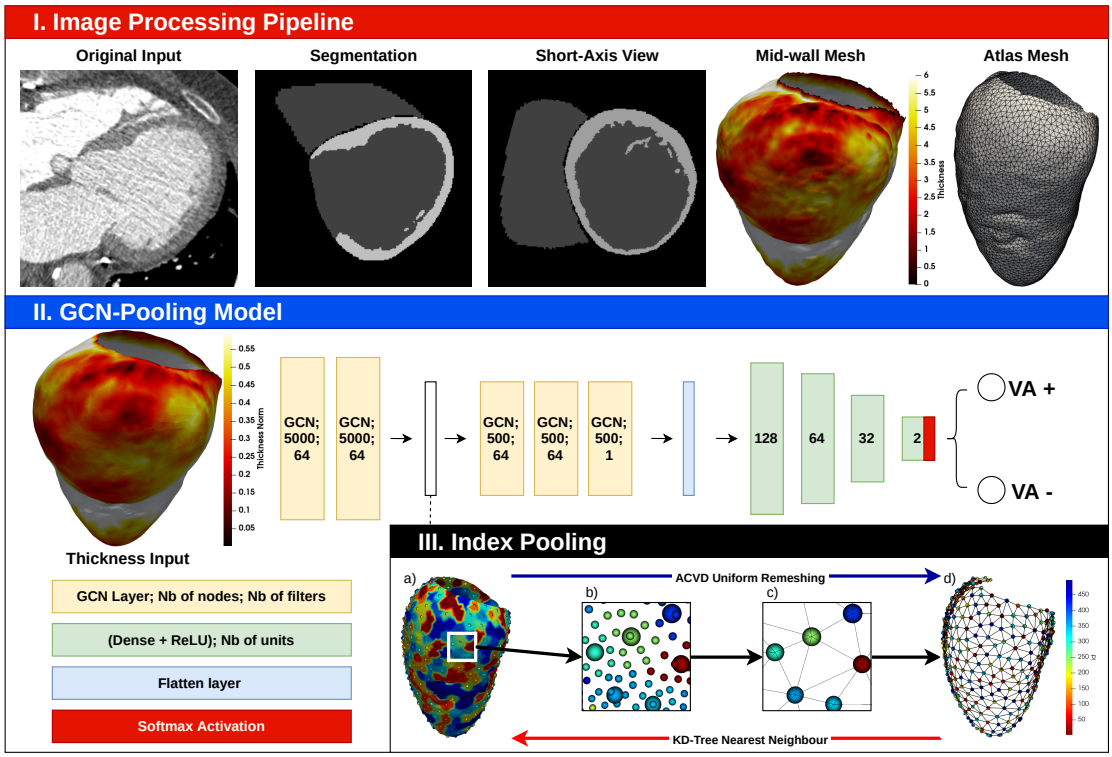


Fig. 5.6.: Image processing pipeline, GCN-Pooling architecture and Index Pooling layer. The thickness input mesh was the result obtained after fitting the average mesh to the mid-wall mesh, ensuring the point-to-point correspondence of every input. *Index Pooling:* a). pooling patches (from KD-tree NN); b). original (smaller spheres) and pooling nodes (bigger spheres); c & d). coarse graph.

[1] https://github.com/danielegrattarola/spektral
[2] https://www.tensorflow.org/

5.2.3 Interpretability Study

We studied the interpretability of the graph neural network by adapting to our context two existing methods: Local Interpretable Model-agnostic Explanation (LIME, [Ribeiro, 2016]), and integrated gradients [Sundararajan, 2017].

5.2.3.1 LIME

LIME or Local Interpretable Model-agnostic Explaination as proposed by [Ribeiro, 2016] is a method to calculate the coefficient $\hat{\beta}_n$, the "interpretable coefficients", associated to each superpixel in the input, corresponding to the importance of the superpixel i toward the model prediction. Briefly, in the case for the image data, LIME computes a local surrogate linear model based on the absence or presence of the superpixels. Our use-case is actually similar if we consider the input graph $\mathcal{G} = (X, A)$ as an image with pixel values defined by the thickness. Following [Garreau, 2021], our method can be described as:

1. segment the mesh into d patches;

2. randomly *turn on/off* the patches to create new inputs:
   $\mathcal{G}_1 = (X_1, A)$,..., $\mathcal{G}_n = (X_n, A)$;

3. run the model prediction on all the new inputs $y_i = f(\mathcal{G}_i)$;

4. compute $\hat{\beta}_n$ by fitting the y_is to a local weighted surrogate model.

The proposed surrogate model to calculate $\hat{\beta}_n$ follows the weighted ridge regression between the interpretable features z_i as input and the model responses $y := f(\mathcal{G}_i)$. The interpretable features z_i denotes the activation ($z_{i,j} = 1$) or deactivation ($z_{i,j} = 0$) of the patches $j \in \{1, ..., d\}$ such that $X_i = z_{i,j}X + (1 - z_{i,j})\bar{X}$. The weights applied to z_i are defined by the cosine distance d_{cos} from the $z = \mathbf{1} = (1, ..., 1)^T$ of the original input X, as following

$$\pi := \exp\left(\frac{-d_{cos}(\mathbf{1}, z_i)^2}{2\nu^2}\right), \tag{5.6}$$

where $\nu > 0$ denotes the *bandwidth parameter*, by default $\nu = 0.25$.

We used two segmentation methods in step 1. The first is *quick shift* [Vedaldi, 2008], as is default for the image version of LIME, which bases the segmentation on both the nodes 3D coordinates and the thickness value[3]. We scaled the magnitude of the node

[3] `https://github.com/Nick-Ol/MedoidShift-and-QuickShift`

coordinates and thickness value using scikit-learn standard feature scaler[4]. The quickshift algorithm was then applied with the threshold of 0.2. The second is *grid segmentation* with KD-tree NN, which is based only on the 3D coordinates of the nodes. The mesh was first downsampled to a 30 mm radius using 3D voxel grid downsampling, before running the KD-Tree search which generated a quasi-uniform grid segmentation of the input mesh.

Turning the patches i *on* meant replacing the original value of patch with the replacement value, and vice versa. To study the case of VA+ classification, the replacement value was set to 1 to represent the normalised healthy LVMYO, under the hypothesis that a completely healthy LV would be classified as VA-. The number of random patch activations was set to 1000. Figure 5.7 shows the example of the segmentation and the patch activation used as inputs in the LIME method.

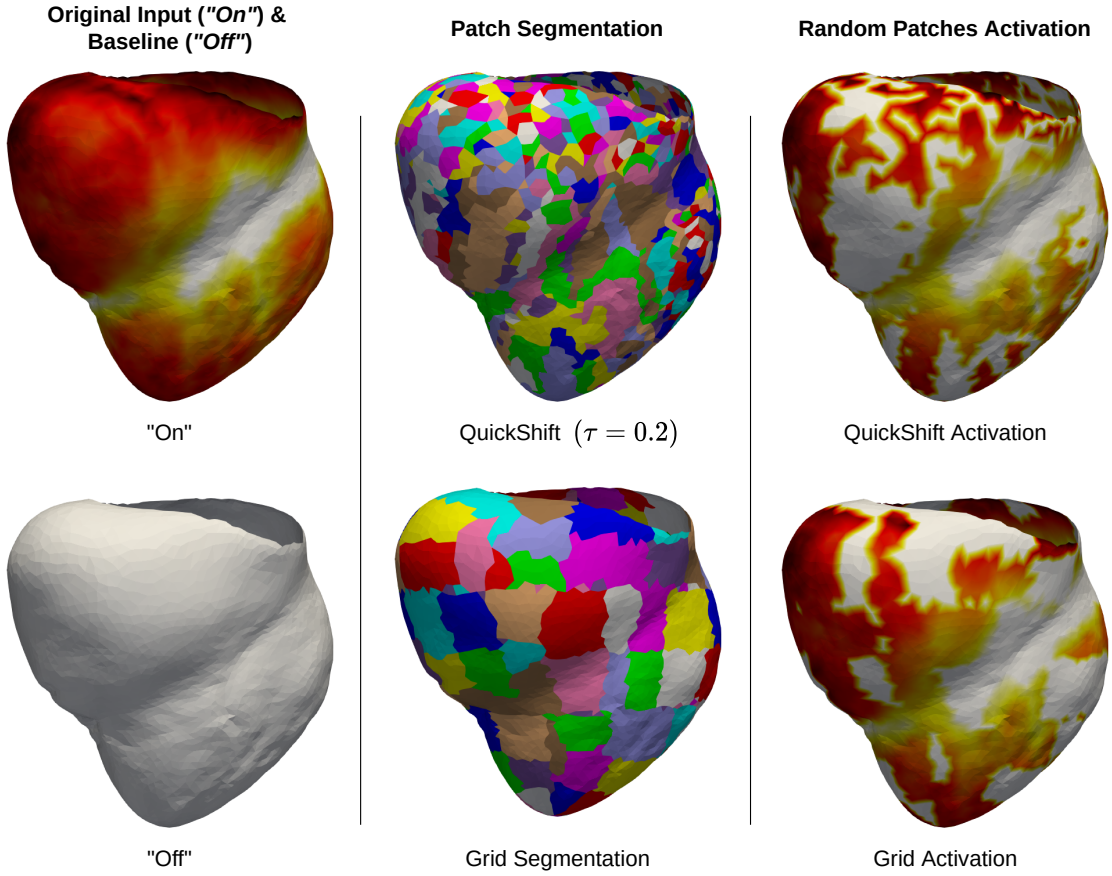


Fig. 5.7.: Mesh input processing for LIME methods. The original input is segmented into smaller patches. Then, the patches are randomly turned *on/off* in order to create new input for each iteration.

[4] `https://scikit-learn.org/`

5.2.3.2 Integrated Gradients

Integrated gradients are computed as the integrals of backpropagated gradients of the straight-line path from the baseline input x' to the original input x, as shown in figure 5.8. To adapt the method to the graph network f, the integrated gradients along the dimension i^{th} for a model $f(\mathcal{G})$ with the input graph $\mathcal{G} = (X, A)$ and baseline graph $\mathcal{G}' = (X', A)$ was calculated as:

$$\mathrm{IG}_i(\mathcal{G} = (X, A)) ::= (X_i - X_i') \times \int_{\alpha=0}^{1} \frac{\partial f(X' + \alpha(X - X'), A)}{\partial X_i} d\alpha \,, \tag{5.7}$$

where $\alpha \in [0, 1]$ denotes the step coefficients between the baseline input ($\alpha = 0$) to the input ($\alpha = 1$). We set the number of steps to 50. We set the baseline feature X' the same way as the replacement input in LIME by changing the thickness feature value to 1.

Behind the theory of the integrated gradients, the model explanation or "attribution" method must satisfy two main conditions: *sensitivity* and *implementation invariance*, respectively.

The *sensitivity* condition imposes that if the inputs differ in a single feature resulting in different predictions, then the feature should be assigned a non-zero attribution, and that if the model does not depend on some feature, then the attribution of that feature should be zero. Following the fundamental theorem of calculus of line integrals, i.e. for the differentiable function φ over the continuous curve between q and p, then $\int_p^q \nabla\varphi(r) \cdot dr = \varphi(q) - \varphi(p)$, the Eq. 5.7 can then be rewritten as:

$$\sum_{i}^{n} IG_i(x) = f(x) - f(x') \tag{5.8}$$

Thus, the output attributions can be viewed as the difference between the output of the model at the original input x and the baseline x', which satisfies the $sensitivity$ condition. This also means that all continuous paths from baseline to input, $P_{i=1,2,3}$, in figure 5.8 can be used. The straight-line path is chosen for its ease of calculation, as well as the symmetry-preserving aspect of the processing between the input and the baseline, as detailed in the Appendix A of [Sundararajan, 2017].

The *implementation invariance* imposes that the attributions are identical for all functionally equivalent models. These are implied to the models that produce the same inputs and outputs, while having different implementations. For example, the attributions from two trained (and converged) models with different set of weights should not differ. This condition is satisfied as the method based these attributions only on the model's gradients.

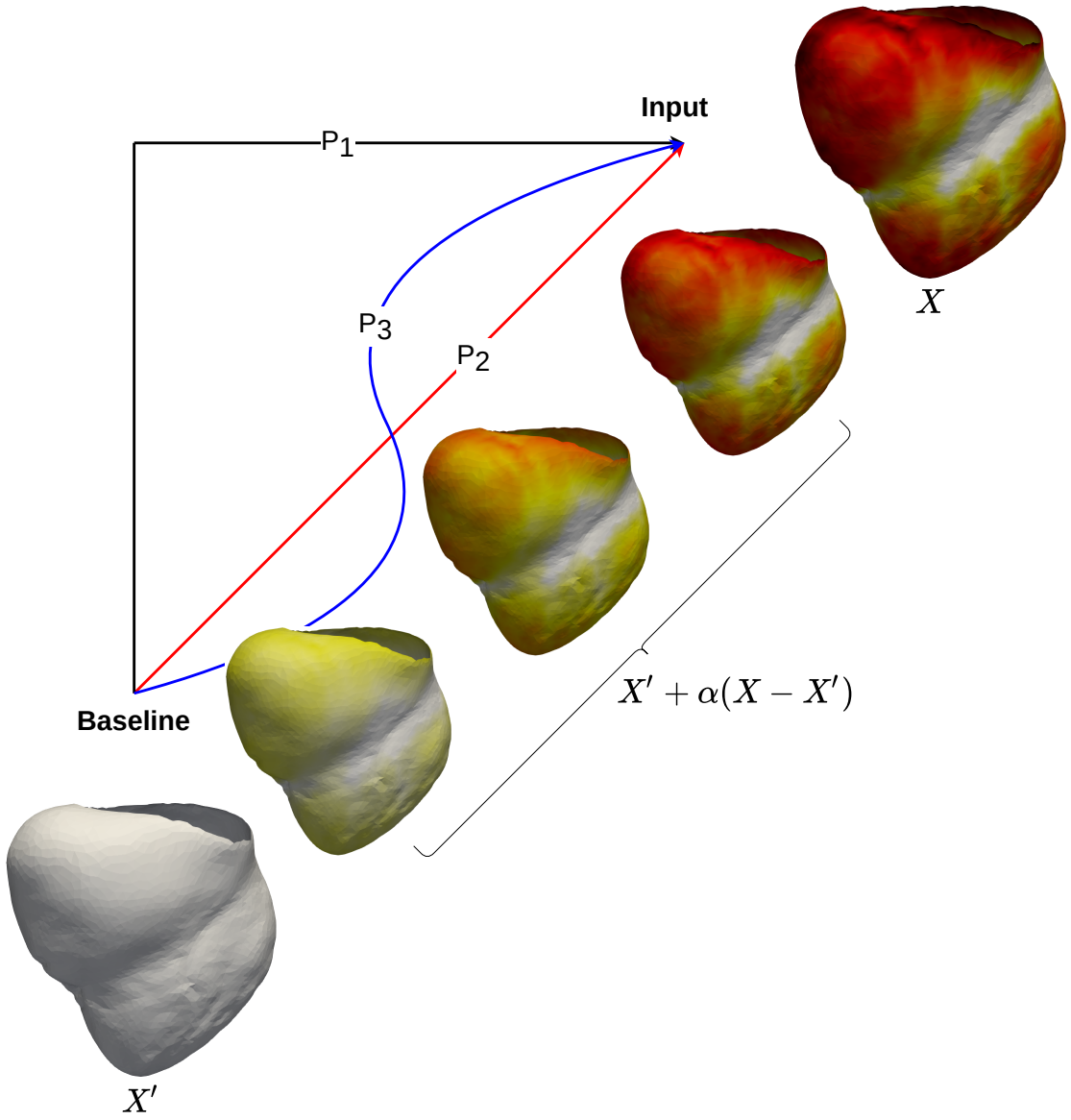


Fig. 5.8.: The linear interpolation example of the Integrated Gradients path (P_2). $P_{i=1,2,3}$ denotes the possible paths that interpolate the between the baseline and the original input.

5.3 Experimental Setup

In order to test the proposed method, we used the same retrospective dataset detailed in the previous chapter 4. We used the training and testing population division as done in section 4.4. The training population was used to determine the best graph model configuration, as well as to compare the novel pipeline with the method described in the previous chapter 4. To this end, the DL models were trained and validated using the cross-validation method.

Finally, the best model configuration was retrained using the available training dataset and compared to the clinical baseline markers calculated in section 4.4.

5.3.1 Deep Learning Baselines Comparison

5.3.1.1 Baseline Models Setup

We studied the performance of the GCN-Pooling model against the baseline DL models using a 10-fold cross-validation on the training dataset. The baseline models were set up to answer the following questions:

1. What is the best graph architecture configuration?

    - To determine if each step of the proposed GCN-Pooling model is effective in the classification task, we run the cross-validation of the graph input using three models: the GCN-Pooling, GCN and FCN model.

    - The GCN model had the same number of the GCN layers as the GCN-Pooling model, without the index pooling layer. The number of nodes is thus consistence between the original input and the flatten input to the FCN block.

    - The FCN model had the same number of layer composition as the FCN block of the two previous models. Running this model would allow us to estimate whether the GCN block contributes the classification accuracy.

    - Finally, to determine the best number of input nodes N, we run the training with two graph datasets: at $N = 500$ and $N = 5000$. The pooling nodes for both datasets were set to $N = 250$ and $N = 500$, respectively.

2. Is the Atlas fitting necessary ?

- To assess the additional value of the point-to-point correspondence of the input dataset, we re-ran the GCN and FCN models on the graph dataset before the atlas fitting.

- To account for the physical node position, each model was re-run twice using the node feature $X \in \mathbb{R}^{N\times 1}$ of the thickness value and $X_{with\_pos} \in \mathbb{R}^{N\times 4}$ of the thickness value and the 3D coordinates of the node.

3. How are the graph models compared to the CNN models ?

   - The graph models were compared to the 2D and 3D CNN models.

   - The 2D and 3D datasets were used to run the CVAE-Class and CNN-Class (the straightforward image classifier, detailed in the previous chapter).

   - For the 2D dataset, we used the LV bullseye thickness from our previous work at 256×256 resolution.

   - For the 3D dataset, we used the SAX-oriented LV wall masks with uniform rotation and slices resolution. Next, the 3D images were cropped around the mask to create a 3D dataset of $240 \times 240 \times 288$ resolution. Finally, the binary mask was used instead of the LV thickness value image, because the model failed to converge with the later input during the trail run of both the CVAE-Class and the CNN-Class model.

The summary of the baseline model and input configurations is shown in figure 5.9.

5.3.1.2 Cross-validation Setup

We used a 10-fold random sub-sampling cross-validation to test the baseline models. The training population (451) was randomly sampled with replacement. Here the key difference from the bootstrap method to create 10 sets of training and testing datasets at 9:1 ratio. To train each model, the training sets were further split into training and validation data, at 8:2 ratio. Each model was optimised and validated, with the initial learning rate at $1e-4$, which was reduced by 0.5 every 5 epochs if no validation improvement was observed. The training was stopped after 50 epochs when there was no validation improvement. The model achieving the best validation score was then used to evaluate the testing set. Finally, the evaluation results from each of the 10 sets were summarised to access the model's performance.

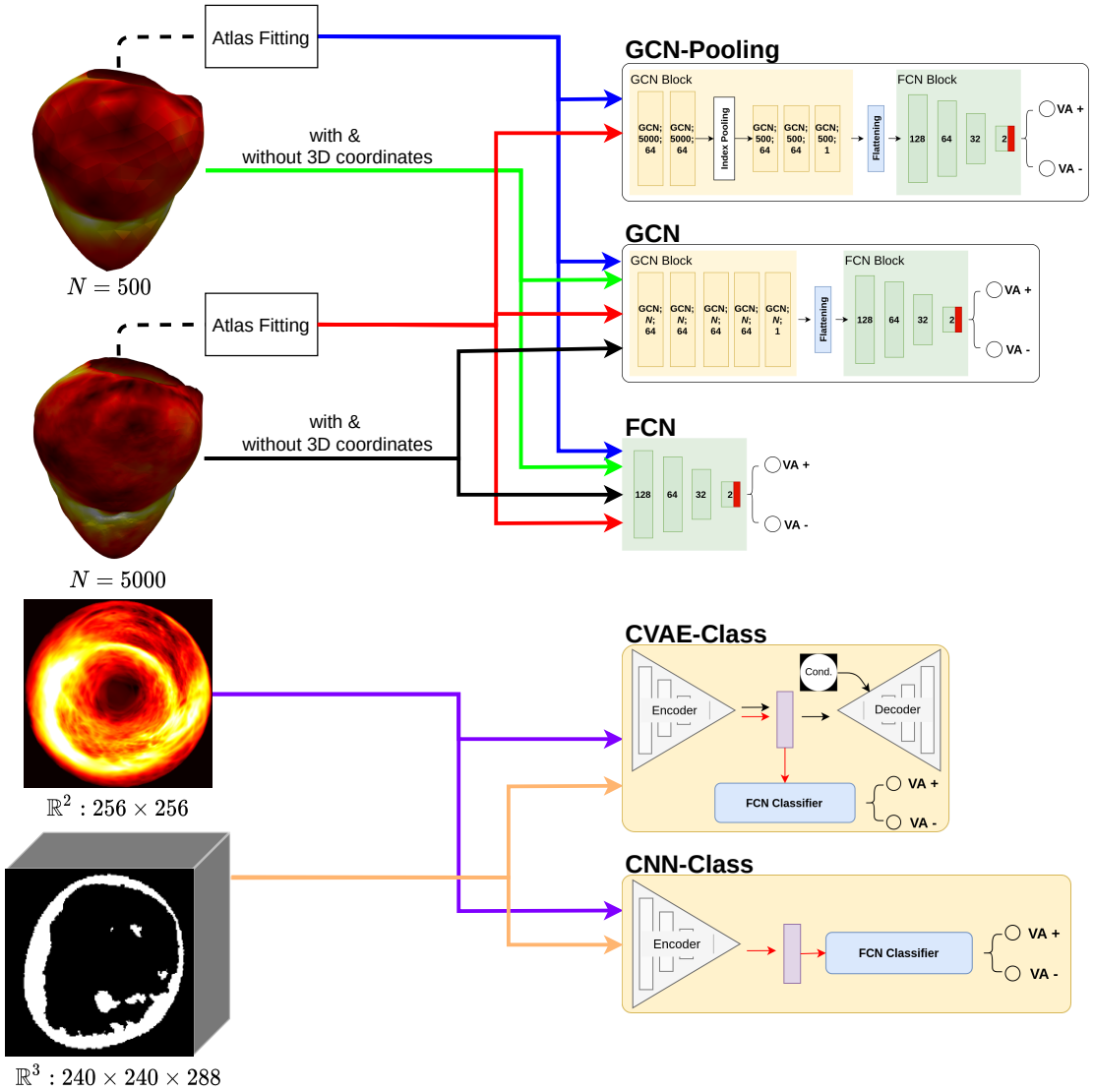


Fig. 5.9.: The input types and their corresponding classification models run in the cross-validation. The training population (451) was randomly sampled into 10 sets, with the training sets and evaluation set at a 9:1 ratio. Each model was trained until an early stopping occurred at 50 iterations without validation improvement. The evaluation was then done using the model weights, which achieved the best validation score during training.

5.4 Results

5.4.1 Evaluation against DL models

Table 5.1 and Figure 5.10 show the cross-validation results of the DL models. Overall, the best performance model was the GCN-Pooling model at $N = 5000$ nodes at 0.818 ± 0.06 accuracy, followed by the 2D CVAE-Class model at 0.788 ± 0.06.

With point-to-point correspondence, the GCN-Pooling model outperformed both GCN and FCN model, whose best models only achieved 0.784 ± 0.05 and 0.770 ± 0.06, respectively. Without the index pooling layer the GCN and FCN performed better with $N = 500$ than $N = 5000$, thus proving that the index pooling is necessary for larger nodes input. The GCN block also improved the accuracy, as both GCN and GCN-Pooling scored higher accuracy than the FCN models.

Without the atlas fitting, the accuracy was significantly reduced with the best model, and the FCN-5000 with positions only reached 0.743 ± 0.03 accuracy. This legitimised the point-to-point correspondence as an indispensable step in the VA classification pipeline.

Both CVAE-Class and CNN-Class models were shown to perform better with the 2D LV bullseye input compared to the 3D input. Furthermore, the CVAE-Class model also showed higher accuracy than the CNN-Class model for both input types. These observations further support our image processing approach and the model architecture proposed in the previous chapter.

5.4.2 Evaluation against Clinical Baseline

We retrained the GCN-Pooling model using the entire training population (451) and evaluated the model using the same testing population as in chapter 4. The evaluation results including the GCN-Pooling model on the testing dataset are shown in Tab. 5.2. The 5000-node GCN-Pooling model achieved the highest prediction accuracy at 0.812, followed by the CVAE-Class at 0.792 and the scar-based markers at 0.725 (scar size) and 0.689 (scar age). The LVEF model had the lowest performance at 0.671 accuracy.

5.4.3 Model Interpretability

Figure 5.11 shows the output coefficients of LIME and integrated gradients on the true positive predictions from the testing population as compared to the GradCAM++ method used in the previous chapter. We could observe the consistency of the high coefficient

Tab. 5.1.: 10-fold cross-validation of the DL baseline models. The results were displayed as $mean$ ($\pm std$). (*) Non average fitted input; (**) Non average fitted input with position and thickness features. **Bold row**: most accurate model.

Model	Input Shape	Accuracy	Sensitivity	Specificity
CNN-class	256×256	0.775 ($\pm$0.05)	0.616 ($\pm$0.10)	0.834 ($\pm$0.06)
	$240 \times 240 \times 288$	0.759 ($\pm$0.06)	0.508 ($\pm$0.19)	0.853 ($\pm$0.05)
(C)VAE-Class	256×256	0.788 ($\pm$0.06)	0.691 ($\pm$0.18)	0.825 ($\pm$0.05)
	$240 \times 240 \times 288$	0.775 ($\pm$0.06)	0.457 ($\pm$0.18)	0.875 ($\pm$0.03)
FCN	$N = 500$*	0.604 ($\pm$0.04)	0.525 ($\pm$0.15)	0.643 ($\pm$0.09)
	$N = 500$**	0.700 ($\pm$0.04)	0.525 ($\pm$0.15)	0.643 ($\pm$0.09)
	$N = 500$	0.770 ($\pm$0.06)	0.716 ($\pm$0.18)	0.790 ($\pm$0.07)
	$N = 5000$*	0.706 ($\pm$0.07)	0.250 ($\pm$0.11)	0.878 ($\pm$0.09)
	$N = 5000$**	0.743 ($\pm$0.03)	0.120 ($\pm$0.09)	0.975 ($\pm$0.03)
	$N = 5000$	0.761 ($\pm$0.12)	0.775 ($\pm$0.17)	0.756 ($\pm$0.17)
GCN	$N = 500$*	0.701 ($\pm$0.05)	0.626 ($\pm$0.09)	0.742 ($\pm$0.11)
	$N = 500$**	0.718 ($\pm$0.07)	0.766 ($\pm$0.17)	0.700 ($\pm$0.09)
	$N = 500$	0.784 ($\pm$0.08)	0.716 ($\pm$0.17)	0.809 ($\pm$0.09)
	$N = 5000$*	0.679 ($\pm$0.12)	0.633 ($\pm$0.17)	0.696 ($\pm$0.21)
	$N = 5000$**	0.677 ($\pm$0.09)	0.566 ($\pm$0.20)	0.719 ($\pm$0.16)
	$N = 5000$	0.779 ($\pm$0.08)	0.716 ($\pm$0.19)	0.803 ($\pm$0.10)
GCN-Pooling	$N = 500$	0.784 ($\pm$0.05)	0.808 ($\pm$0.07)	0.775 ($\pm$0.06)
	$\boldsymbol{N = 5000}$	**0.818** ($\pm$**0.06**)	**0.766** ($\pm$**0.10**)	**0.837** ($\pm$**0.07**)

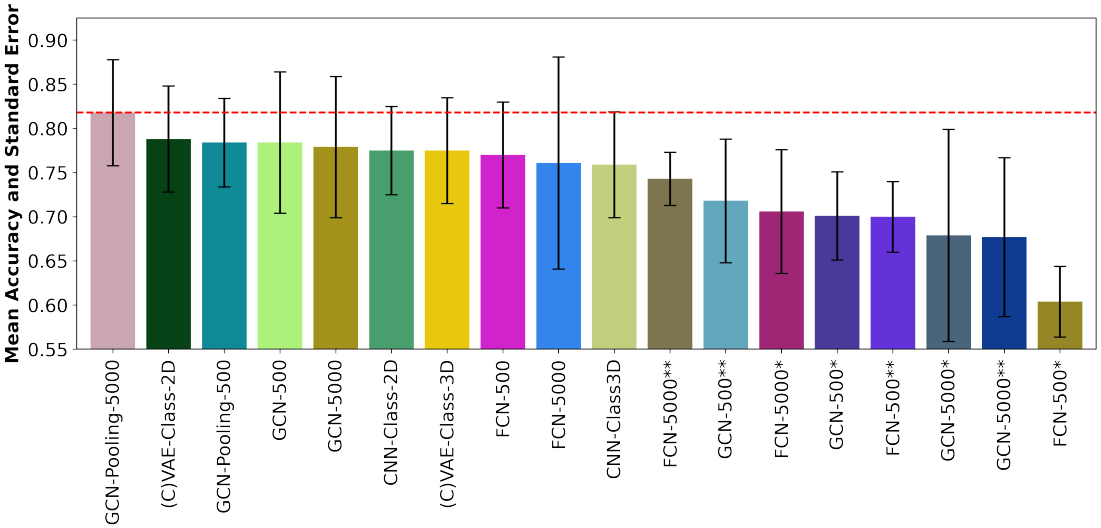


Fig. 5.10.: 10-fold cross-validation of the DL baseline models. The results were sorted out by the mean accuracy and standard error, in descending and ascending order, respectively. (*) Non average fitted input; (**) Non average fitted input with position and thickness features. The red dotted line indicates the mean accuracy of the GCN-Pooling model with 5000 nodes.

Tab. 5.2.: Evaluation results on the testing population of the GCN-Pooling model and the clinical baselines.

	Accuracy	Sensitivity	Specificity
GCN-Pooling ($N = 5000$)	**0.812**	**0.780**	**0.824**
CVAE-Class	**0.792**	**0.732**	**0.815**
Scar Size $> 65.49cm^2$	0.725	0.683	0.741
Scar Age $> 132months$	0.698	0.839	0.640
LVEF $< 43\%$	0.671	0.707	0.657

(blue) regions between the two methods. The high coefficient regions were also faithful among LIME outputs when the different segmentation methods were used.

We also noticed that the regions with prominent coefficient correlated strongly with the thinning regions of the LVMYO. This validates the hypothesis that the post-infarction LV thinning is associated with the presence of VA. Moreover, a closer inspection of the top coefficient patches showed that the GCN-Pooling model did not base its VA+ prediction on the entire thinning regions, as opposed to the GradCAM++ method. Instead, the higher coefficients were rather focused on smaller distinctive regions within or adjacent to the scar regions, as indicated by the yellow ellipsoids in figure 5.11.

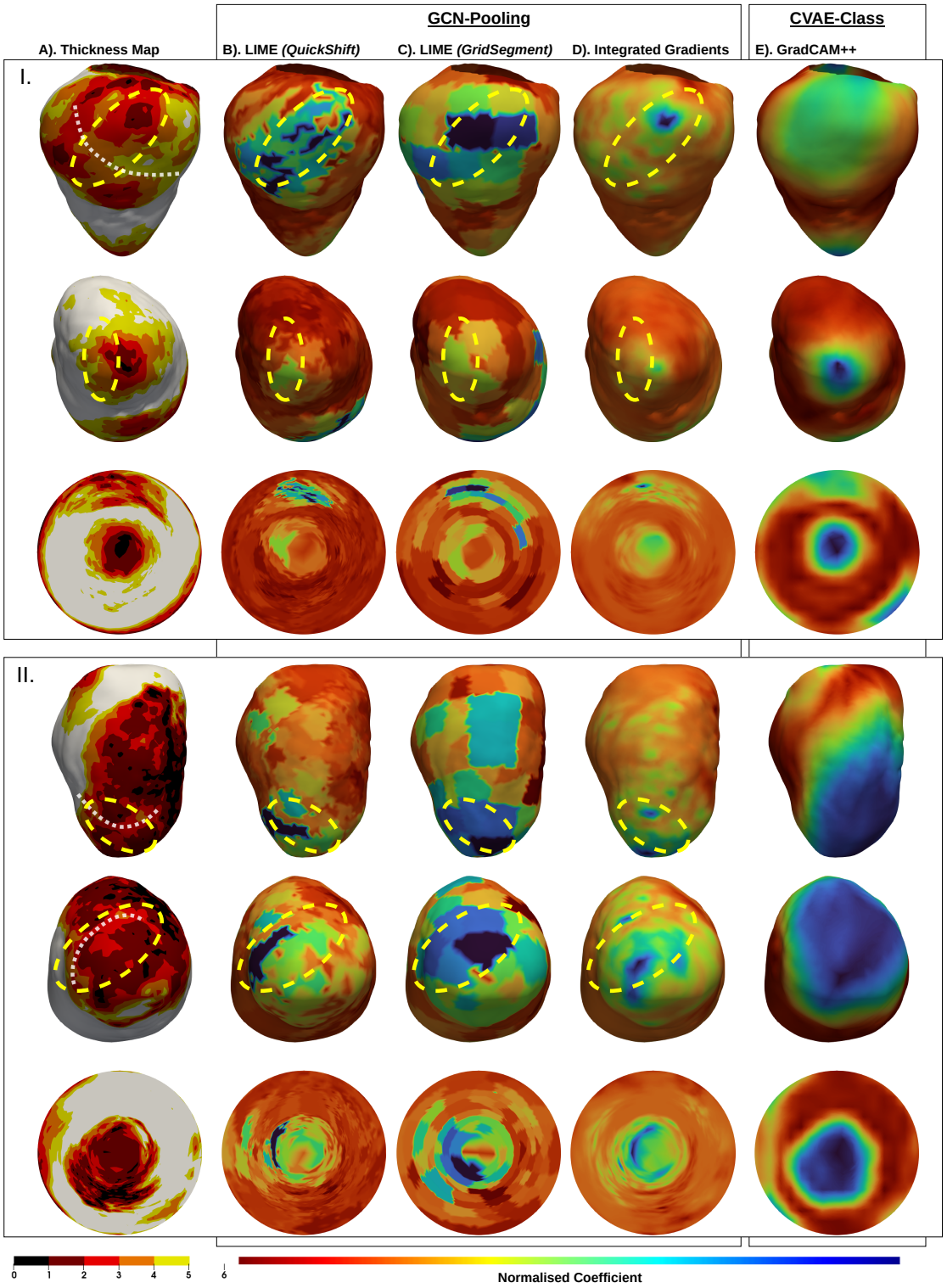


Fig. 5.11.: Interpretability studies of the GCN-Pooling model on two true positive predictions from testing population. The coefficients from LIME and integrated gradients are more focused on the distinctive regions within the thinning area, compared to the GradCAM++ method where the entire area is highlighted. The yellow ellipsoids highlight the regions with a consistently high coefficient among different methods. The white dotted curves highlight the potential reentry channel (i.e., the substrate of VA). First row: inferior LV view; second row: apex view; third row: the bullseye map projection.

5.5 Discussion

5.5.1 VA classification with GCN-Pooling.

With the current input to graph model (GCN-Pooling, $N = 5000$) $\mathcal{G}(X, A)$ the only geometry information fed to the model is in the adjacency matrix A without any other indicator of the node position and relation in the physical space. During the model development, we assumed that this relationship was already embedded via the point-to-point correspondence. Nonetheless, explicit position features might provide the model with extra-information for more robust classification.

Notably, the GCN layer used to build the model does not allow the inclusion of the edge feature. This layer was chosen for its ease of implementation, making it one of the most popular layers. However, more complex graph layers [Gong, 2018; Simonovsky, 2017; Jiang, 2020], where the edge features were integrated along with the node feature, were shown to improve the classification accuracy. For instance, the distances between the nodes can be integrated as an edge feature in order to take into account the edge distortions resulted from the atlas fitting, as seen in Figure 5.12. While the original meshes were uniformly re-meshed (which resulted in equi-distance edges), the affine registration of the atlas introduced heterogeneous edge distortions. In this particular case, the distance information could be useful for the model to correctly measure the extent of the thinning region.

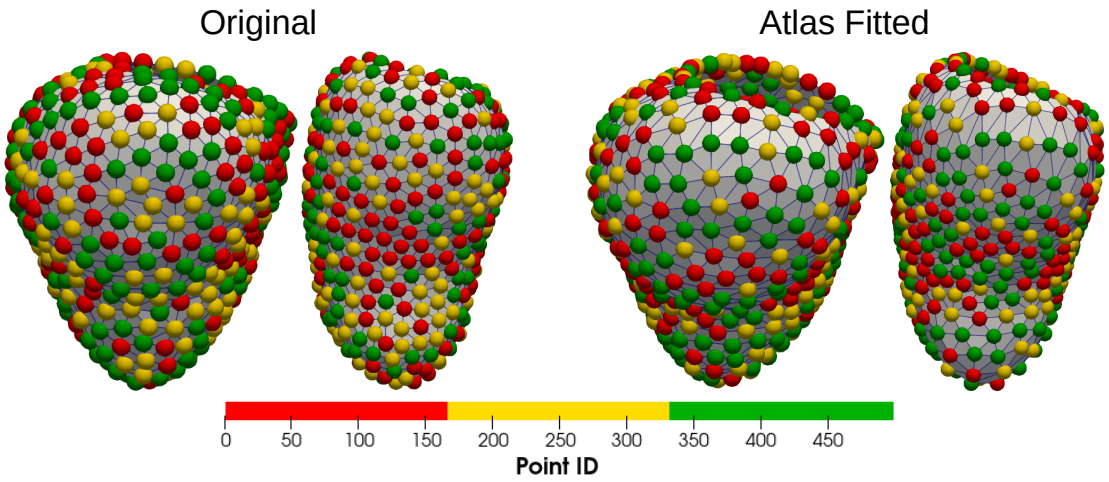


Fig. 5.12.: Comparison between original and atlas fitted meshes. The nodes were colour coded according to their ids. The point-to-point correspondence can be observed in the uniform pattern of the nodes between the atlas fitted meshes. We can also observe the none uniform edges length between the nodes, resulted from the affine transformation of the atlas mesh.

5.5.2 Interpretability Study

The interpretability analysis only provided a perspective into the relation between the model prediction and the input. The high coefficient regions might be highly arrhythmogenic, simply because these were the regions allowing the model to classify the input as VA+. While the choice of the segmentation method in LIME is not trivial, we focused the analysis only on two segmentation methods, to limit the scope of this study. The quick shift method was chosen since it incorporates both physical coordinates and the thickness values of the node, and was parameterised to generate small segments (by setting smaller threshold). In contrast, the grid segmentation method does not consider the thickness value and was parameterised to generate larger segments (using larger neighbour searching radius). Therefore, the consistency of the two LIME methods, as well as the integrated gradients, further solidify the tolerance of our interpretability approaches towards the superpixel segmentation bias.

Likewise, the choice of the baseline for the two methods is not trivial. Both methods were proposed as computer vision classification tasks, where the target usually exhibits very distinctive pattern. Substituting the original input with a unique value would likely change the output prediction. In this study, we set the baseline features to $X = \mathbf{1}$, which would lead to the VA- prediction. While this would not disagree with the integrated gradients method, we might unintentionally introduce new reentry channels. In the LIME method, such patches of fast conducting healthy myocardium within the scar region, as can be observed in Figure 5.7. However, the large number of random patch resamplings and the agreement with the integrated gradients method could allow a degree of confidence for the current LIME configuration. Moreover, although the existence of the reentry channel and its relation to VA+ class in the dataset would ultimately influence the model optimisation, it can be rather far-fetched to claim that the classification model bases its prediction directly and solely on the presence of a reentry channel.

Nevertheless, limiting by the retrospective aspect of the available data, the current analysis of the interpretability outputs were restricted to a hypothesis.

5.5.3 Study Dataset

Although the testing dataset was not used during the model designing and tuning, the cross-validation and evaluation results were still based on the mono-centre dataset. The retrospective nature of our dataset could also introduce additional bias related to the specific protocol and guideline of the imaging centre, which is dependent on a selected population. Therefore, the current study population might not be a true representation of the general public.

Furthermore, we suggest that a prospective and multi-centre dataset is necessary to confirm the generalisability of the prediction pipeline and the validity of the detected arrhythmogenic site.

5.6 Conclusion

In this work we proposed an improved automatic pipeline for sacr-related VA predictions using myocardial wall thickness defined from cardiac CT images. Formulating the LV thickness map into a graph input allowed the model to perform an accurate prediction of VA based on the 3D anatomy of the LV, achieving a better score than the model trained on the 2D input. The evaluation on the testing dataset showed that the GCN-Pooling model outperformed the clinical markers, especially compared to the current gold standard predictor (i.e., clinical LVEF index). Our interpretability studies proved that the proposed model derived its prediction from the very specific regions within larger scar regions. This encouraging result is important for our future work which will focus on personalised identification of arrhythmogenic site of the scar-related VA within the left ventricle.

Conclusions

6

Contents

In this chapter, we summarise the contributions of the thesis. Lastly, considering the current limitations of our developed methods, we suggest some perspective methodological and clinical studies for future work to validate the application of cardiac CT in VA classification and to localise arrhythmogenic sites, as well as to advance this research towards a robust implementation of multi-modal data for VA prediction.

6.1 Contributions

6.1.1 Methodological Contributions

6.1.1.1 Image Processing

Concerning the image processing tasks, the main contributions of this thesis are directly related to the processing of the 3D cardiac CT images as the input dataset to the classification DL model.

- In chapter 3, we proposed a pipeline to transform raw cardiac CT images into a more suitable data representation for DL model training. Specifically, we established automatic image processing steps to generate a 2D bullseye representation of the LV wall thickness from CT images.

- The LV mesh generated in section 3.2.3 can be used to extract the size of the scar region and can potentially be used in the clinical practice to visualise the geometry and morphology of LV.

- In section 3.2.4, we propose an automatic method to select the flattening reference points to ensure a uniform rotation of the input, as well as its agreement with the AHA 17-segment model.

- In section 5.2.1, we propose a fully automatic method to generate the atlas model of the LV surface mesh, which allows us to generate a point-to-point correspondent graph input dataset, which is suitable for the graph network training, as demonstrated in 5.4.

6.1.1.2 Deep learning

The main contribution of our work related to DL can be split into the two task studies: segmentation and classification. For the segmentation task, our works contributed to improving the segmentation accuracy of the MR images (chapter 2), with data augmentation as well as the anatomical constrains of the cardiac structures.

For the classification task, we investigated the utility of both convolutional (chapter 4) and graph convolutional model (chapter 5) in order to the perform accurate VA classification. We also proposed the adaptation of the existing explainable and interpretable methods to solve the "black box" issue, as well as to leverage the data-driven method to characterise the features of scar regions, potentially arrhythmogenic.

- In section 2.2, we have demonstrated that using a multiple input modality in the training dataset, including the synthetically modified images from the proposed style data augmentation, could improve the segmentation accuracy of an unseen image modality. This work also illustrates the possibility in training a generalised segmentation model capable of segmenting any image modality, simply by curating the appropriate training dataset.

- In sections 2.3.1.1 and 2.3.1.2, we proposed additional anatomical constrains to improve the segmentation of the scar regions from the LGE MR images. Here we proposed an automatic method to dynamically adjust the weights of the different loss functions in section 2.3.1.3. We then demonstrated in section 2.3.2, that the integration of these methods improved significantly the segmentation accuracy of the scar region.

- In section 4.2.1, we proposed the conditional variational autoencoder classifier (CVAE-Class), which was proven to achieve higher VA classification accuracy than the direct CNN classification model.

- In section 5.2.2, we formulated the LV surface mesh into a graph input and proposed a GCN-Pooling model, which produced a better classification accuracy than the prediction pipeline based on the traditional convolutional-layer model.

- We investigated the utility of explainability method, GradCAM++, to provide transparency to the model prediction, as detailed in section 4.2.2.

- In the next section 5.2.3, we adapted two interpretability methods (LIME and integrated gradient, respectively) and used them with the GCN input and model, which allowed us to generate more refined and specific regions of arrhythmogenic sites, as seen by the classification model.

6.1.2 Clinical Contributions

The main clinical contribution of this thesis is the development of a robust CT image-based pipeline that can be rapidly translated into clinical routine to identify LV wall thinning and use it for VA classification and SDA death risk stratification. The methodological works in the chapter 4 and 5 showed that the automatic classification of VA from CT images is feasible and allowed more accurate classification than the conventional clinical markers.

- The evaluation setup and the results obtained in section 4.4 proved the validity of the scar characteristics derived from CT imaging for the VA classification, and this method is superior to the clinical markers such as the current gold standard predictor, the LVEF. We also proved that the use of the DL model (i.e., CVAE-Class) could further improve the classification accuracy.

- The explainability (section 4.5) and the interpretability (section 5.4.3) results validated the LV wall thinning role in arrhythmogenesis. Moreover, the encouraging results proved the intepretability study to be a potential data-driven method for the automatic localisation of the VA substrate.

6.2 Publication

The following publications were the direct results of the works done in this thesis.

- Buntheng Ly, Hubert Cochet, Maxime Sermesant. Style Data Augmentation for Robust Segmentation of Multi-Modality Cardiac MRI. STACOM 2019 - 10th Workhop

on Statistical Atlases and Computational Modelling of the Heart, Oct 2019, Shenzhen, China. URL: https://hal.inria.fr/hal-02401643v3/document

- Buntheng Ly, Sonny Finsterbach, Marta Nuñez-Garcia, Hubert Cochet, Maxime Sermesant. Scar-Related Ventricular Arrhythmia Prediction from Imaging Using Explainable Deep Learning. FIMH 2021 - 11th International Conference on Functional Imaging and Modeling of the Heart, Jun 2021, Stanford, United States. URL: https://hal.inria.fr/hal-03378951/document

- Buntheng Ly, Sonny Finsterbach, Marta Nuñez-Garcia, Pierre Jais, Damien Garreau, Hubert Cochet, Maxime Sermesant. Interpretable Prediction of Post-Infarct Ventricular Arrhythmia using Graph Convolutional Network. STACOM 2022 - 13th Workhop on Statistical Atlases and Computational Modelling of the Heart, Sep 2022, Singapore, Singapore. https://hal.inria.fr/hal-03829609/document

The VA classification pipeline with the CVAE-Class model was implemented as part of the AI pipeline in the MUSIC software, which provides the visualisation support for each image processing steps. The implementation work was described as part of the following publication:

- Merle M, Collot F, Castelneau J, Migerditichan P, Juhoor M, Ly B, Ozenne V, Quesson B, Zemzemi N, Coudière Y, Jaïs P, Cochet H, Sermesant M. MUSIC: Cardiac Imaging, Modelling and Visualisation Software for Diagnosis and Therapy. Applied Sciences. 2022; 12(12):6145. URL: https://www.mdpi.com/2076-3417/12/12/6145/pdf?version=1655454174

The CVAE and CVAE-Class model architecture were also successfully implemented in two collaborative projects for the ECG imaging (EGCI) reconstruction and eye fundus diagnosis. The ECGI is a non-invasive imaging technique that generates cardiac activation maps using the body surface electrical signals and the cardiac geometry. In this collaboration, we successfully employed the 3D version of the CVAE model, where the conditional inputs are the cardiac geometry and the body surface signals, to produce the corresponding cardiac activation maps. For our collaboration in the eye fundus diagnosis project, the CVAE-Class model was evaluated on the eye fundus spectral images and was proven to achieve a robust classification to accurately differentiate the healthy cased from diseased patients.

These collaborative works resulted in the following publications:

- Tania Bacoyannis, Julian Krebs, Nicolas Cedilnik, Hubert Cochet, Maxime Sermesant. Deep Learning Formulation of ECGI for Data-driven Integration of Spa-

tiotemporal Correlations and Imaging Information. FIMH 2019 - 10th International Conference on Functional Imaging and Modeling of the Heart, Jun 2019, Bordeaux, France. pp.20-28. URL: https://hal.inria.fr/hal-02108958/document

- Francisco Javier Burgos-Fernandez, Buntheng Ly, Fernando Díaz-Doutón, Meritxell Vilaseca, Jaume Pujol, Maxime Sermesant; Deep learning for eye fundus diagnosis based on multispectral imaging. Invest. Ophthalmol. Vis. Sci. 2022;63(7):2098 – F0087.

During the thesis, I was a part of the authors of a book chapter on the disease outcome prediction using AI in cardiology. The chapter was included in the following book:

- (2022) AI and Big Data in Cardiology: A Practical Guide (1st ed.). Springer International Publishing.

6.3 Conclusion

To sum up, in this manuscript we proved that it is feasible to automatically process non-invasive imaging data for a robust VA classification. Our work based on the CT imaging further emphasised the relation of the LV thinning with the myocardial scar and post-infarct VA. The statistical evaluation in section 4.4 demonstrated that the LV thinning alone can be used for a better classification than the gold standard clinical marker, the LVEF. Additionally, using the DL models proposed in chapter 4 and 5, we showed that the LV thickness heterogeneity can be analysed with the automatic models to achieve an even more robust classification.

Finally, using the explainable learning methods, we introduced the transparency needed for the DL model to be accepted in the clinical context. The results in chapter 5 also demonstrated the potential of data-driven methods to localise the VA substrate sites.

6.4 Perspective

Our future works will focus on the improvement of the data-driven classification of VA and identification of the arrhythmogenic sites.

The classification models proposed in our work only used the LV wall thickness as input. The goal of our study was to validate the relation between LV wall thinning and post-infarct VA, which we have successfully proven by the evaluation results to be a better indicator than the clinical markers. However, it might be beneficial to build a

VA classification model that could leverage both the LV wall thickness and the clinical markers in order to improve the current classification accuracy. This direction should be explored particularly since the current dataset can be rapidly employed.

The inclusion criteria for VA in the current classification task does not distinguish between ventricular tachycardia, ventricular fibrillation, or aborted cardiac arrest. Although these episodes can be referred under VA, their mechanisms are inherently different. However, it is imperative to distinguish the prediction classes between these episodes. This would not only lead to a more interesting clinical application, but it can also potentially improve the accuracy and viability of the explainable learning output.

To bring the current method closer to clinical relevancy, we also need to test our methods using a prospective dataset. The validation method can be split into two stages: the VA classification and the localisation of arrhythmogenic site via interpretability study. While the first validation would require a clinical trail with mid to long-term follow-up, the second validation could be more conceivable. First, the method can be compared with the automatic method, which can reproduce VT activation patterns recorded during an RFA procedure, as proposed by [Cedilnik, 2020]. This would allow us to exploit the current retrospective dataset. Secondly, as the pipeline is automatic the localisation method can be easily tested directly on the images acquired during the RFA planning and the arrhythmia patterns recorded during the interventions.

While the CT imaging has its advantages against the LGE CMR imaging, the latter still remains the gold standard imaging technique for the scar regions. Through CMR imaging, the information inaccessible by CT imaging such as the transmurality and density of the myocardial scar can be extracted, which could be the key to improve the classification of the VA. Nonetheless, the integration of the CMR images into our automatic pipeline necessitates an accurate segmentation of heterogeneous myocardium in different regions (i.e., dense scar, substrate, healthy tissue), which would ultimately require a larger training dataset if the DL model is to be used.

Lastly, under the current EHRA/AHA guidelines, the imaging data was not given a major role in the SDA or VA patients management. However, we advocate that the VA classification pipeline using the non-invasive data and the localisation of arrhythmogenic sites can be used in clinical practice to support a better patient selection, a superior therapy planning and a more effective RFA guidance.

Appendix

A

Novel Pipeline to Predict VA Risk using Personalised Simulation: Validation using a Pig Model of Chronic Infarct

This chapter contains preliminary work relevant to ventricular arrhythmia risk predictions using personalised virtual models built for in silico electrophysiology simulations. Such virtual models can complement the information obtained using the deep-learning methods presented in the previous chapters. However, virtual models still need to be properly validated in order to be translated into clinical routine diagnostic of scar-related VA. The models presented in here were constructed using preclinical MR images and catheter-based mapping/VA data obtained in preclinical studies (data shared through our collaboration with Sunnybrook, Toronto). Note that we only focused on presenting the analysis pipeline along with preliminary results, reason for which this section was included as an annexe.

A.1 Introduction

As mentioned in previous chapters, scar-related ventricular arrhythmia is one of the most important causes of death worldwide. Typically, for anatomical re-entry, the substrate of VT is caused by the heterogeneous nature of the infarcted tissue. The infarcted tissue includes dense, non-conductive scars, as well as a mixture of viable myocytes and collagen fibres that is located at the interface between dense scars and healthy myocardium. Often, the arrhythmogenic substrate forms slow conductive channels (i.e. isthmuses) that protrude into non-conductive scars. Such geometrical and electrical circuits facilitate the action potential waves to propagate through these channels and loop around the scars. Thus, the VT cycle length (i.e. time to loop around the scars) is shorter than a normal rhythm generated by the Sinus Node (SN) and depends on the scars size, as well as on the physical properties of the critical isthmus and the healthy tissue. Such properties include: the electrical conductivity (which tunes the speed of AP wave) and the anisotropy (i.e. fibre directions). The smaller the scar, the faster and more dangerous the VT cycle length, although the electrical properties and circuit morphology play important roles in determining how fast the VT wave will loop around the scars.

Furthermore, as already mentioned in the previous chapters, the precise localisation of arrhythmogenic sites harboured within the scars along with the predictions of scar-related VT risks are critical problems in the EP cath-lab.

In the past years, computational heart modelling has been established as a powerful tool capable to predict via in silico simulations the VT inducibility for scar-related geometries using preclinical and clinical heart-specific cases. Specifically, for such virtual predictions, one needs to first build detailed 3D heart models that integrate imaging-defined tissue zones with different properties, as well as fibre directions. Further, the models need to be coupled the 3D models with mathematical equations and numerical solvers to find the solution for transmembrane voltage (action potential wave) and its propagation thorough the heart. The different structural zones (e.g. dense scar, VT substrate, and healthy tissue) can be identified using CMR; however, the current clinical slice thickness is relatively large (8-10 mm), leading to partial volume effects and inadequate identification of the VT substrate. In addition, it is difficult to measure the fibre directions and integrate them into the 3 models; thus, other solutions should be sought. Lastly, theoretical approaches replicate certain clinical stimulation protocols that are employed in the arrhythmia cath-lab for VT inducibility. However, in order to validate these virtual models, it is important to use accurate model parameters, ideally measured directly during the electrophysiology (EP) mapping studies and VT tests.

In this work, we sought to develop a pipeline for predicting activation maps and scar-related VT inducibility, using a preclinical swine model a chronic infarction that replicates human pathophysiology. The novelty of the pipeline consists in using: an atlas of fibre directions instead of ruled-based generated fibre directions (as in [Cedilnik, 2020]); high-resolution in vivo 3D MR-based biventricular models; and, a fast computational model that would allow us to predict VT inducibility risk and associated VT patterns from many virtual points (including the one that coincided with the stimulation point used in a real EP study, for validation purposes). The MRI-EP datasets were acquired at Sunnybrook Research Institute, Toronto, Canada and shared through our collaboration.

Notably, here we first focus on describing the imaging processing step of generating meshes from segmented MR images followed by the deformation steps required to integrate the fibre directions into these models from a recently developed atlas of cardiac fibres. We also describe how we plan to run the VT simulations in the future work, and how the pipeline will allow us to personalise the electrical conductivity (which tunes the speed of AP wave) in different anatomical zones, from the activation maps recorded during the X-ray guided EP mapping procedure.

The diagram in figure A.1 outlines the different steps of the project.

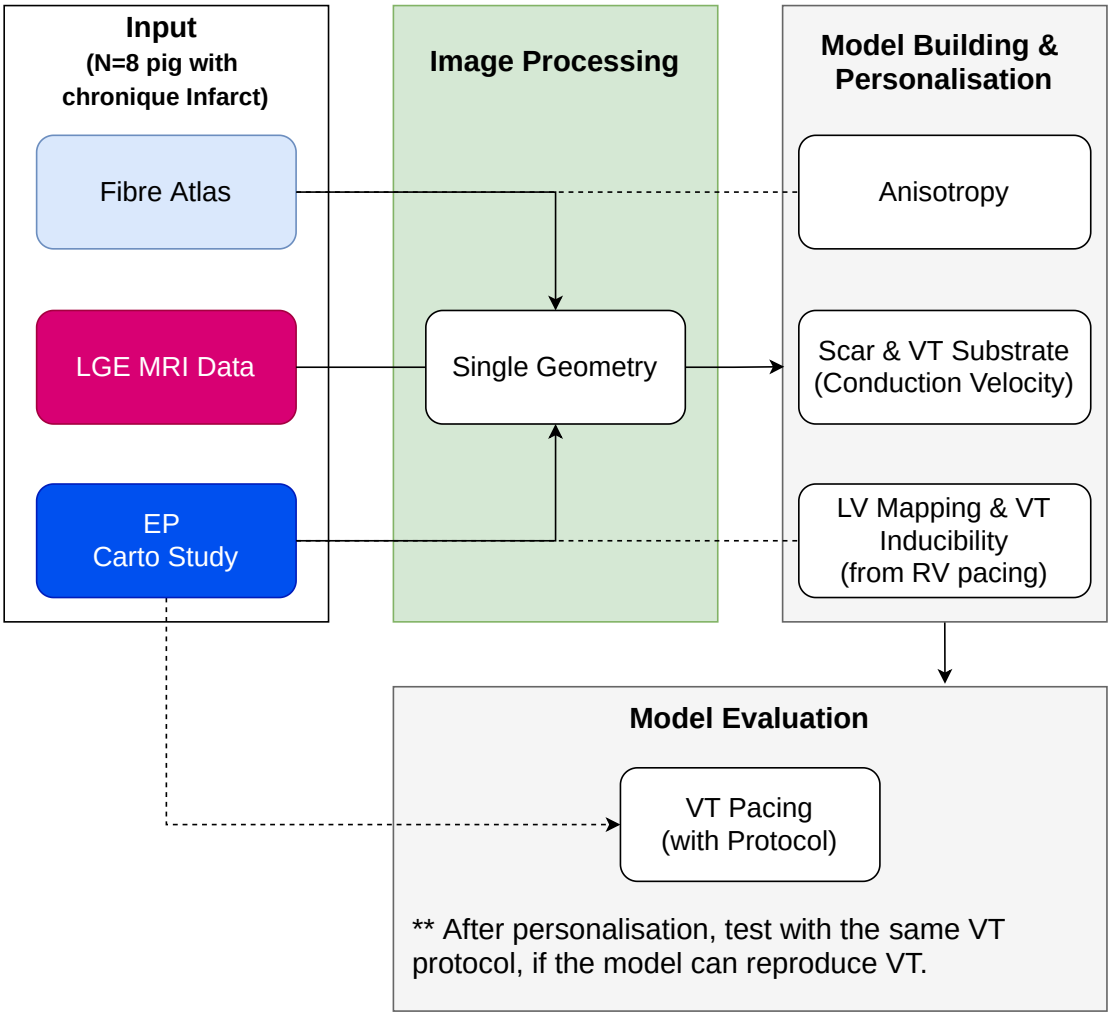


Fig. A.1.: Diagram of the pipeline. The information from different inputs were transferred to each individual heart geometry extracted from the corresponding 3D LGE images. From there, the model building and associated personalisation could be done based on the anisotropy, geometry/location of scar and VT substrate, as well as the information from electrical mapping. Finally, we will compare the activation patterns during pacing and the VT inducibility outcome for each personalised virtual model using the mapping and VT inducibility protocols employed during the real EP studies.
The green box denotes the work that has been already accomplished, while the gray box corresponds to our future work.

A.2 Data Description

For this project, we worked with three types of data, as described below:

- MR imaging data was obtained in n=8 swine with chronic scars, following 5 weeks of healing after the infarct creation. CMR imaging was obtained on a 3T GE scanner, using a free-breathing 3D late gadolinium enhanced (LGE) method employed at 1.4 mm isotropic resolution. The VT substrate in these LGE images can be identified as a 'grey zone', which has an intermediate signal intensity (SI) between healthy

tissue and dense scars. A standard signal intensity threshold of 5SD was used to classify the grey zones (GZ) and scar pixels, which were then used to generate masks with 3 distinct zones: scar, GZ and healthy tissue.

- All pigs underwent an X-ray guided EP study. Electro-anatomical voltage mapping (EAVM) of the left ventricle (LV) was performed using a CARTO3 system (Biosense, Johnson & Johnson). The endocardial LV maps were acquired with Pentaaray catheters, in sinus rhythm and under pacing conditions. Finally, a VT inducibility test was performed using a protocol based on rapid pacing from RV-apex (to override the sinus rhythm), followed by 2 or 3 extra-stimuli, mimicking the clinical inducibility protocols.

- For fibre directions, we used the recent pig atlas developed in a collaborative study [Mojica, 2020]. The atlas was constructed using data obtained from explanted healthy hearts acquired using high-resolution diffusion tensor (DT) imaging methods, at a sub-millimetre spatial resolution.

A.3 Image Processing

A.3.1 Registration of the tensors associated with fibre directions, from the Atlas to a 3D LGE image

Figure A.2 shows the complete processing pipeline that we developed in order to interpolate the fibre tensor orientation from the atlas to an individual LGE geometry. To register the fibre tensor orientation of the atlas to each LGE geometry, we applied an affine registration with Deformetrica, as detailed in chapter 5.

First, we generated the surface mesh of the atlas geometry then match its associated surface mesh to the surface mesh of the LGE mesh, after which the same registration parameters of the surface registration were applied to the point cloud mesh of the atlas. Finally, the tensor matrices of the registered atlas point cloud were interpolated to the point cloud of the LGE image using Log-Euclidean metrics [Arsigny, 2006].

The surface mesh and point cloud mesh of the fibre atlas were generated using the marching cube algorithm from the Atlas mask, thresholded from non-zero tensor matrix. The same processes were then used to generate the surface and point cloud mesh of the LGE image. We next applied a rigid registration of the atlas mesh toward the LGE mesh by matching the centre of mass and the automatically detected positions of the LV and RV.

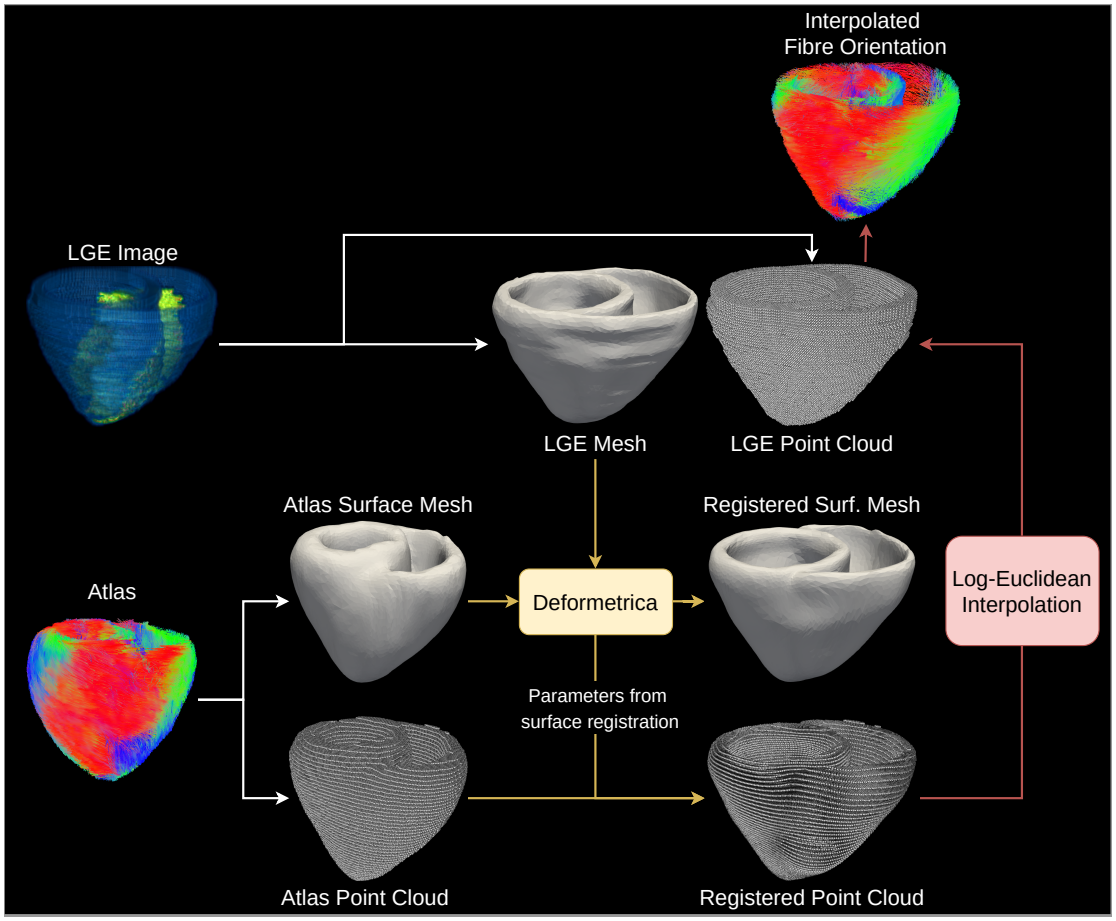


Fig. A.2.: The overview of the fibre orientation registration pipeline. The surface meshes and point clouds were extracted from the Atlas and LGE images. We used Deformetrica to register the surface meshes (i.e., the Atlas mesh to LGE mesh), after which the registration parameters were applied to register the point cloud.

Deformetrica used the grid-based momentum for this registration; therefore, the point cloud mesh could be registered using the same parameters to achieve the same registration. The affine registration outputs are shown in figure A.3.

Finally, the orientation matrices from the registered atlas point cloud to the point cloud of the LGE point cloud, where the orientation values could be easily transferred to the LGE image. Following the method proposed in [Arsigny, 2006], the linear interpolation of the tensors was done in the logarithmic space, by transforming the component of the symmetric matrix, S, using the equation A.1.

$$\log(\mathbf{S}) \simeq \vec{\mathbf{S}} = (\log(\mathbf{S})_{1,1}, \log(\mathbf{S})_{2,2}, \log(\mathbf{S})_{3,3}, \sqrt{2}\log(\mathbf{S})_{1,2}, \sqrt{2}\log(\mathbf{S})_{1,3}, \sqrt{2}\log(\mathbf{S})_{2,3})^T, \tag{A.1}$$

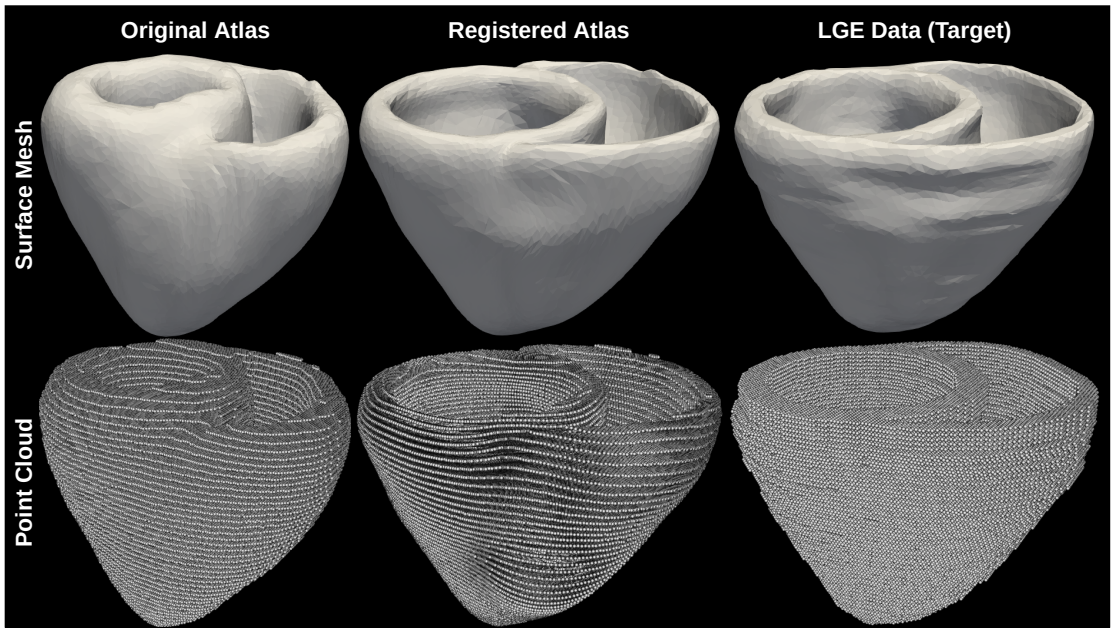


Fig. A.3.: Affine registration outputs. The affine registration of the surface meshes ensures the integration of the ventricular anatomy, while also reduce the number of transformation points, which makes the calculation faster. Then, the point cloud could be registered with the same parameters to achieve the same registration result.

where $\log(\mathbf{S})_{i,j}$ denotes the coefficient of $\log(\mathbf{S})$ in the position (i, j) of the tensor matrix. The log representation of the matrices are then interpolated, using euclidean linear interpolation, and transformed back to the normal scale. At the end of this processing step, we obtained the fibre orientations that correspond to the geometry of a given 3D LGE image, as illustrated in figure A.4.

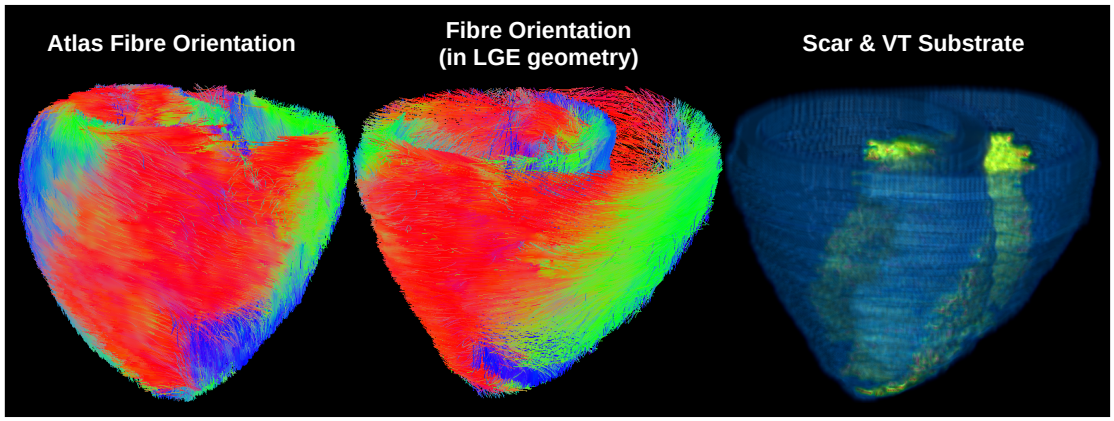


Fig. A.4.: The comparison between the original atlas fibre orientation, the interpolated fibre orientation and the individual LGE image, with the information of the scar and VT substrate shown in green.

A.3.2 Matching EP mesh to LGE mesh

To use the information from the EP study in model personalisation, we designed an automatic method to register the CARTO mesh to the mesh corresponding to the LV endocardium (ENDO) extracted from the LGE image. For this method, we used as the reference mesh the mesh obtained during the RV pacing study. This allowed the automatic detection of the relative position between the LV and RV, which was necessary in order to ensure the correct rotation during the rigid registration step. Figure A.5 details the registration steps performed in this section.

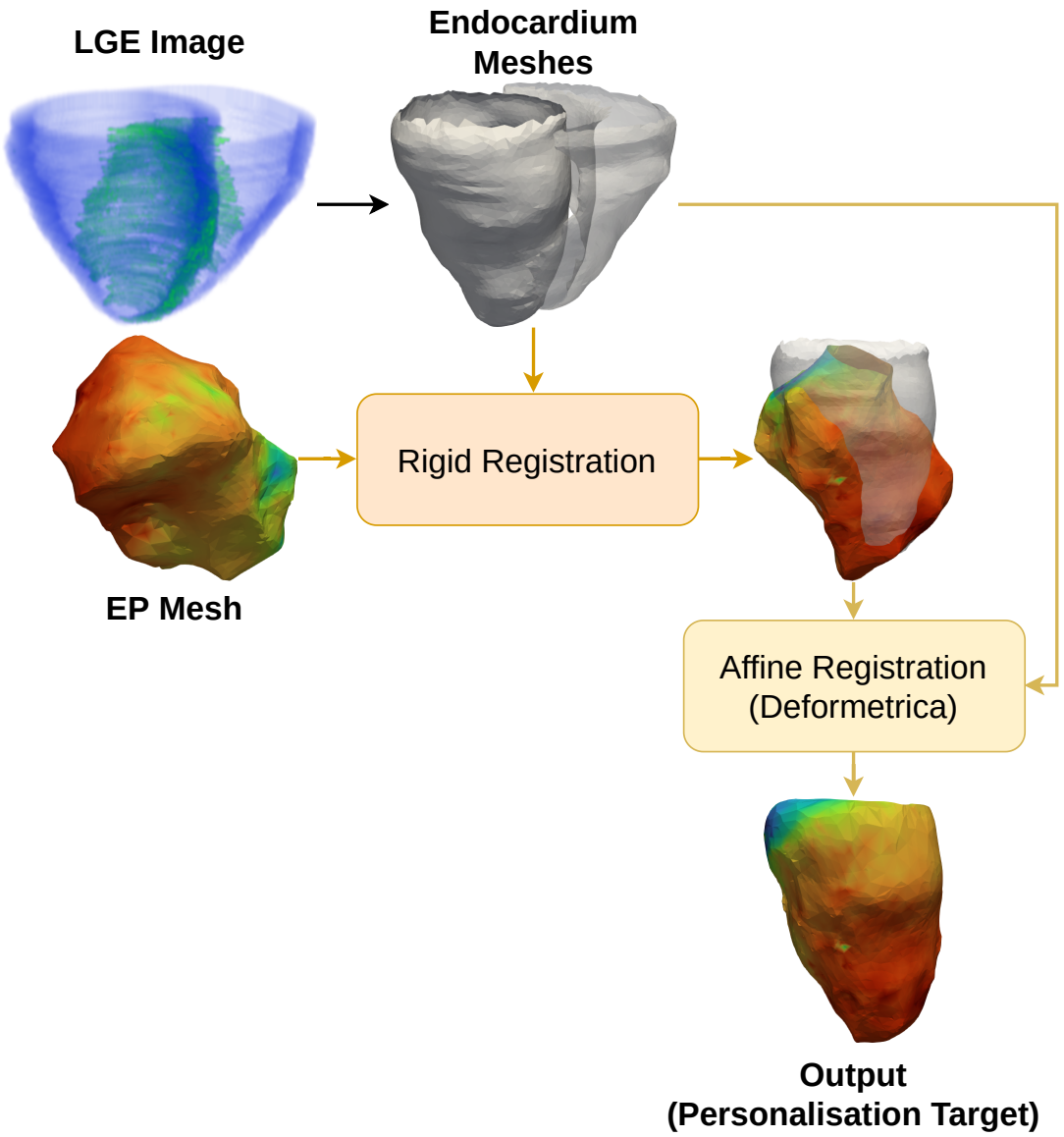


Fig. A.5.: The LV and RV meshes were extracted from the LGE image. The LV mesh was used as the rigid and affine registration target of the input EP mesh.

A.3.2.1 Rigid Registration

We applied rigid registration of the EP mesh to the geometry of the LV ENDO mesh. We based the rotation transformation on two vectors: the LV major axis and the perpendicular axis pointing towards the LV point with earliest activation time. The second vector allowed the alignment of the RV direction of the two meshes. Since the RV position was not usually available in the outputs of the EP study, the corresponding point on the LV was selected as replacement.

On the LGE mesh, the earliest activation point on the LV map was defined as the point on the LV ENDO closest to the RV ENDO apex. Since the LGE images were acquired in 3D, any orientation can be chosen. For simplicity, here we used the short axis (SAX) for the automatic localisation of the apex and basal region, which was done based on the z coordinate.

For the EP mesh, we developed an automatic method to locate the pacing point and the major axis based on the following assumptions.

1. The ENDO mesh is ellipsoidal with a major axis connecting the apex and base of LV.

2. The corresponding pacing point on the LV (in the RV pacing mapping study) is located at the earliest activation point and should be closer to the apex then to the basal region.

To integrate these two assumptions, we used 1-component principal component analysis (PCA), fitted with the activation time and the euclidean position of each point of the mesh. Based on the PCA output, the earliest activation time was selected to be the LV stimulation point. Figure A.6 illustrates the difference between 1-component PCA and the RV pacing activation time map.

To define the major axis of the EP mesh, we fitted the 1-component PCA using the node position. The major axis was defined as the line between the centre of the mass of the 10 percentile regions of both extremities. The axis direction was defined based on the relative position of the end points to the pacing points. To accommodate the ENDO shape, we applied the same PCA method to the LGE mesh. Figure A.7 illustrates the PCA map and the major axis of the EP mesh and the LGE mesh, respectively. The rigid registration was done step-by-step, by aligning the major axes and then the pacing points. Figure A.8 (top row) shows the outputs of the rigid registration.

To sum up, this work represents our first attempt to automate the rigid registration process between an EP study mesh and an LGE-defined mesh. Although the preliminary results seem promising and the registration results are adequate, further testing and

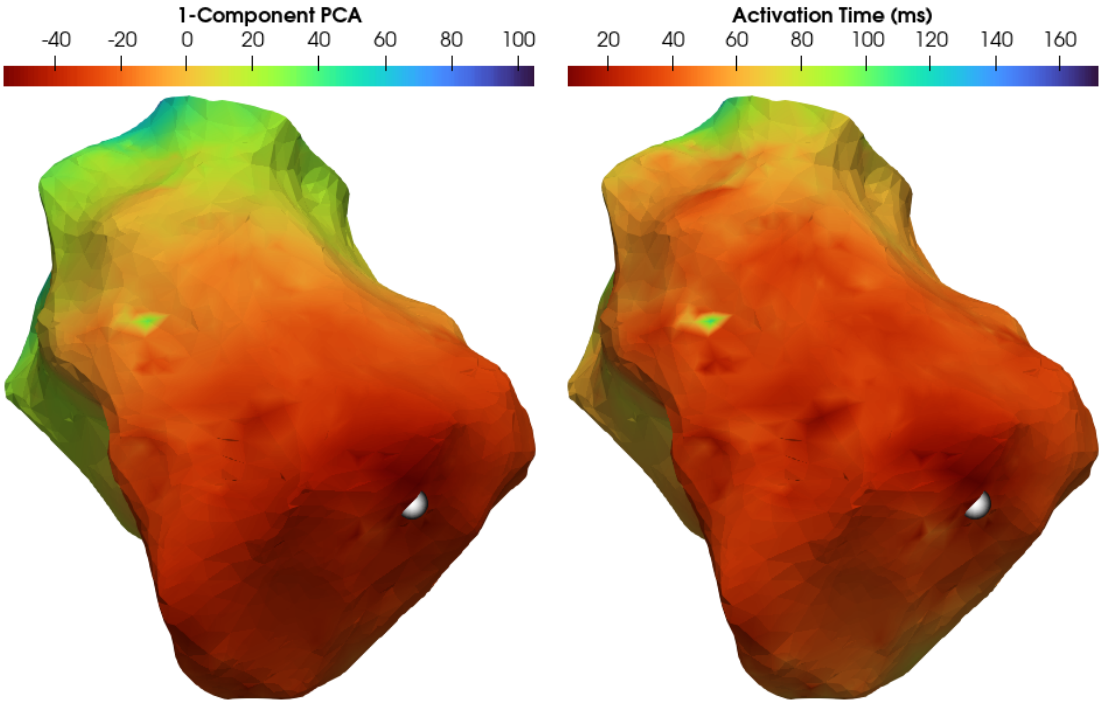


Fig. A.6.: Comparison between 1-component PCA and RV pacing activation map. The addition of the position in the PCA regularised the noise and the global activation map according to the major axis position of each point. The white dot denotes the detected pacing point using the 1-component PCA map.

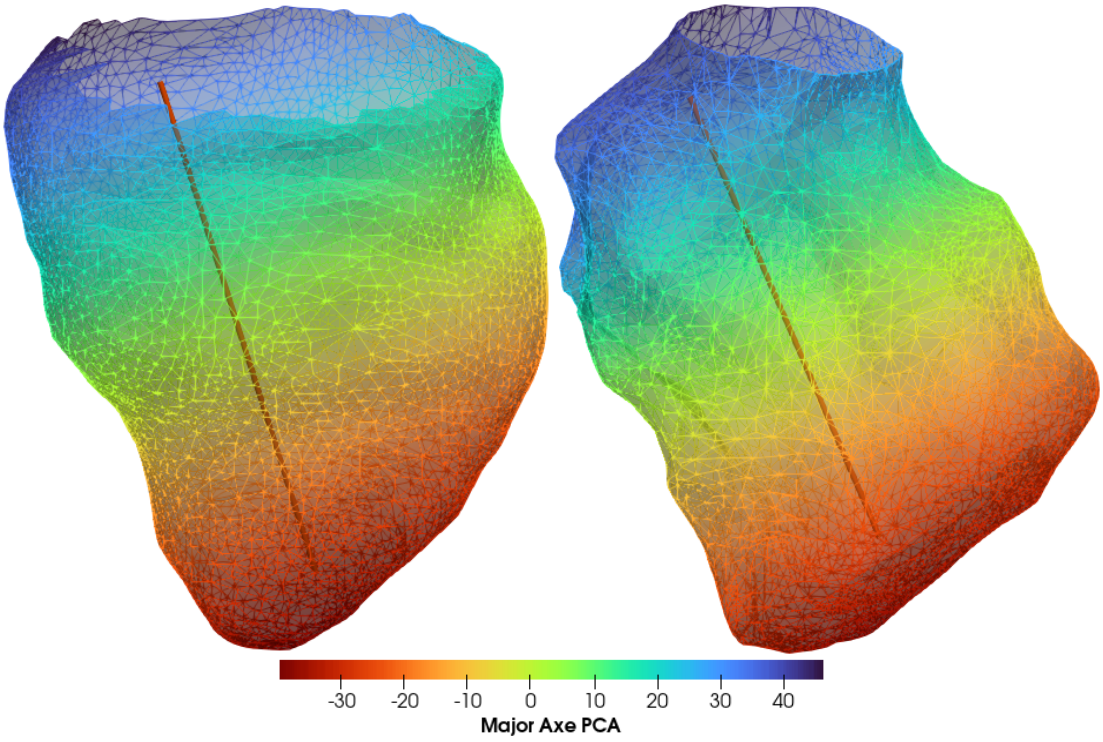


Fig. A.7.: The major axes defined by PCA of the point positions.

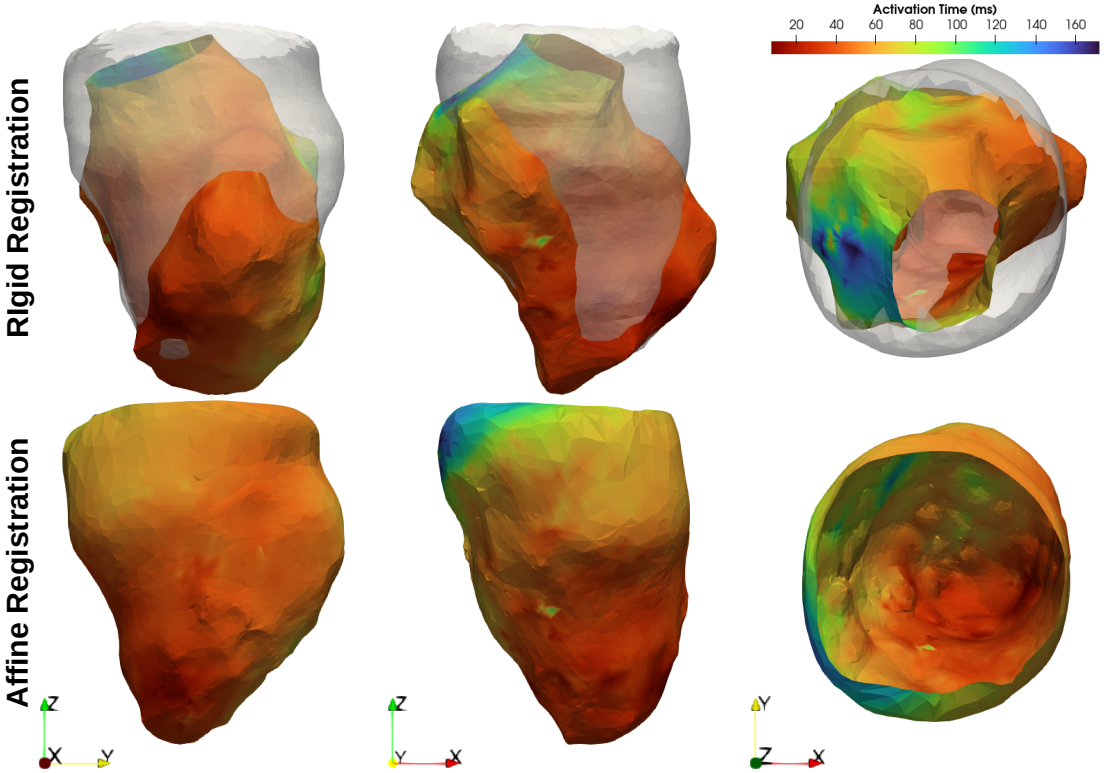


Fig. A.8.: The registration outputs. The rigidly-registered images are superimposed onto the LV ENDO mesh generated from the LGE mask. Top row: rigid registration; Bottom row: affine registration.

evaluations with different qualities of EP meshes are needed in order to finely tune and validate this method.

We must also note that the example EP mesh shown in this section does not reflect the general mesh quality of the EP study. In addition, in the context this project with limited number of cases such as ours, manual rigid registration might be a better suited solution.

A.3.2.2 Affine Registration

After ensuring the rotation alignment with the rigid registration, we applied the affine registration using Deformetrica, as in the previous sections. The output of the affine transformation is shown in figure A.8 (see bottom row).

A.4 Cardiac Electrophysiology Model for Simulations

The model will be built based on the Mitchell-Schaeffer model [Mitchell, 2003] and the lattice Boltzmann Method framework, as recently proposed by our group [Cedilnik,

2020]. Given a proper parameterisation, this theoretical framework could effectively replicate both the depolarisation (phase 0) and the refractory period (phases 0 to 3) of the action potential, as detailed in 1.2.1.3.

The depolarisation in phase 0 can be personalised via the parameter corresponding to electrical conductivity. Thus, different values for this parameter will be assigned to different tissue zones. Figure A.9 illustrates the hypothetical model outputs following the envisioned conductivity personalisation step.

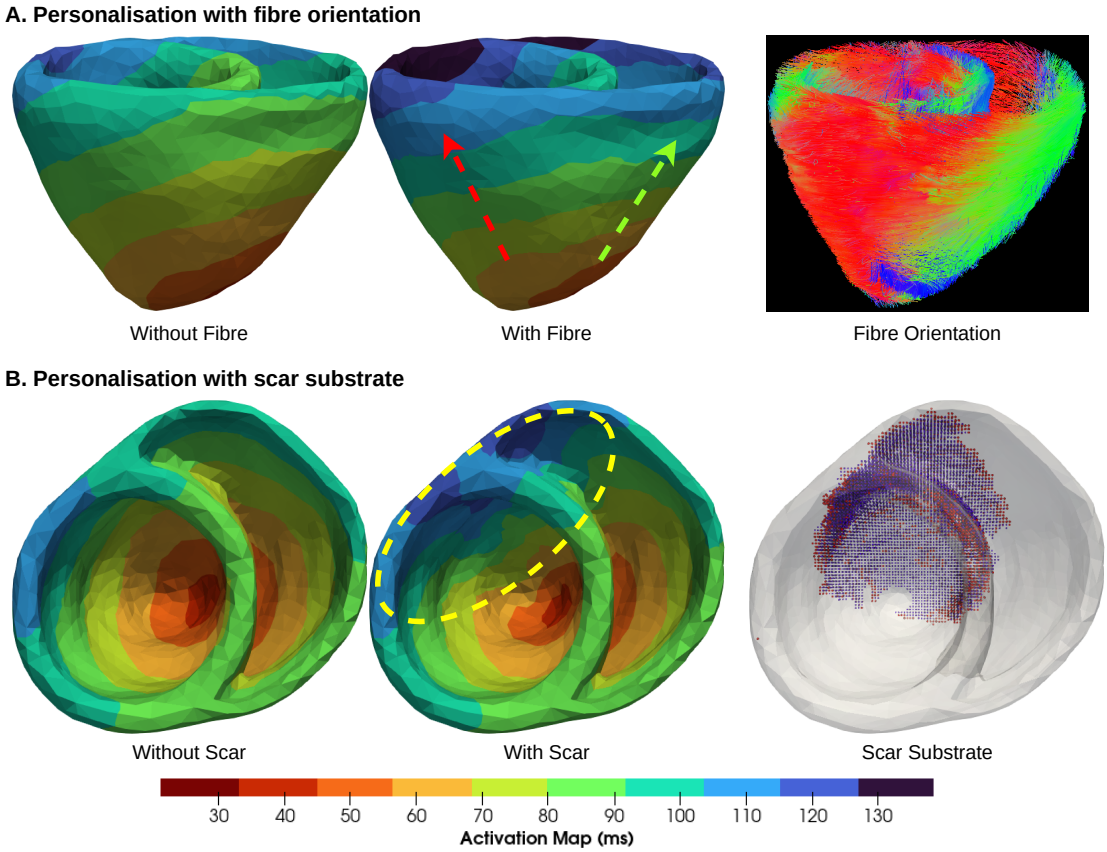


Fig. A.9.: Conductivity personalisation based on fibre orientation (A) and scar/substrate (B). The initial excitation point was set at the apex of the RV.
A. The red arrow indicates the slower propagation in the transverse direction of the fibre, while the green arrow indicates the normal propagation in the parallel direction of the fibre.
B. The yellow ellipse indicates the slower conduction in the potential VT substrate. Within the substrate, the red point cloud indicate the grey zone, while the blue point cloud indicate the non-excitable dense scar.

Our preliminary test using manual adjustment showed that it is possible to closely match the pattern of simulation activation map with the recorded map from the EP study, figure A.10.

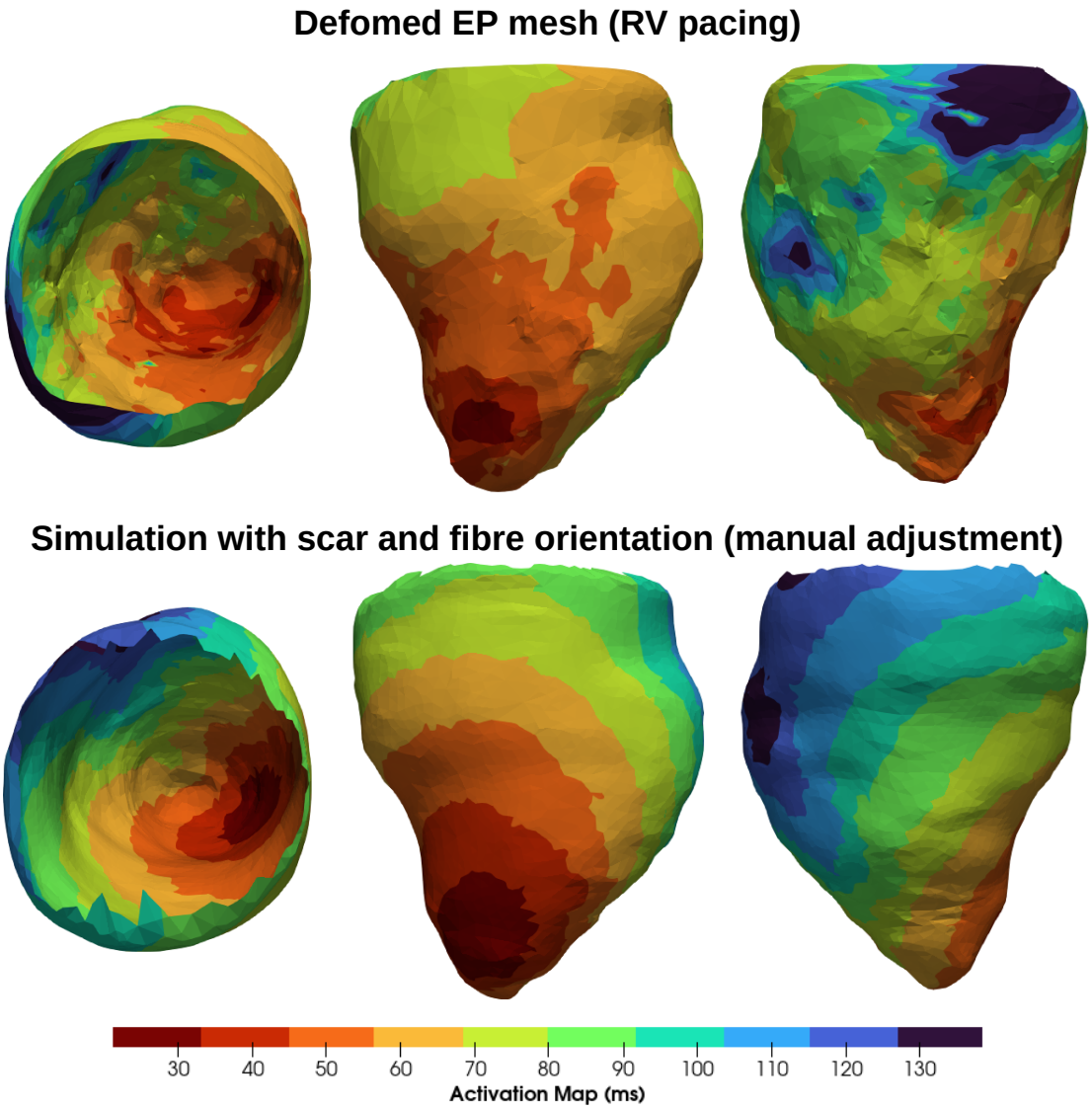


Fig. A.10.: Qualitative comparison between the activation maps recorded (i.e., on the EP mesh) and the simulated activation map produced by the model based on the substrate and fibre orientation parameterisation

A.5 Future Work

In the future work, the following model personalisation and improvement steps will be performed:

- Automatic personalisation of the depolarisation (phase 0) to better match the conductivity with the substrate and fibre orientation.

- Estimation of the fibre remodelling within the gray zone. Note that in the dense scar region the myocardial fibres are replaced by collagen; this zone is not excitable.

- Personalisation of the action potential duration (APD) (phases 0-3) to complete the AP cycle. This could be done by using a clinical surrogate of AP duration, named activation-recovery interval (ARI). ARI can be extracted from the unipolar electrograms recorded during the EP study [Xia, 2005].

- The complete personalisation of the AP cycle would allow us to simulate activation patterns, VT inducibility and its potential re-entry patterns using 101 virtual stimulation points, as in the method proposed by [Cedilnik, 2020], and validate the proposed pipeline.

Bibliography

[, 2022] "2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC)". In: *European Heart Journal* (Aug. 2022). ehac262. eprint: https://academic.oup.com/eurheartj/advance-article-pdf/doi/10.1093/eurheartj/ehac262/45535741/ehac262.pdf (cit. on p. 18).

[Abbet, 2020] Christian Abbet, Inti Zlobec, Behzad Bozorgtabar, and Jean-Philippe Thiran. "Divide-and-Rule: Self-Supervised Learning for Survival Analysis in Colorectal Cancer". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, et al. Cham: Springer International Publishing, 2020 (cit. on p. 71).

[Al-Khatib, 2018] Sana M. Al-Khatib, William G. Stevenson, Michael J. Ackerman, et al. "2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death". In: *Circulation* 138.13 (2018), e272–e391. eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIR.0000000000000549 (cit. on pp. 4, 17–19).

[Anter, 2018] Elad Anter, Andre G. Kleber, Markus Rottmann, et al. "Infarct-Related Ventricular Tachycardia: Redefining the Electrophysiological Substrate of the Isthmus During Sinus Rhythm". In: *JACC: Clinical Electrophysiology* 4.8 (2018), pp. 1033–1048 (cit. on p. 15).

[Arsigny, 2006] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. "Log-Euclidean metrics for fast and simple calculus on diffusion tensors". In: *Magnetic Resonance in Medicine* 56.2 (2006), pp. 411–421. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20965 (cit. on pp. 136, 137).

[Berruezo, 2015] Antonio Berruezo, Juan Fernández-Armenta, David Andreu, et al. "Scar Dechanneling". In: *Circulation: Arrhythmia and Electrophysiology* 8.2 (2015), pp. 326–336. eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIRCEP.114.002386 (cit. on p. 18).

[Biffi, 2018] Carlo Biffi, Ozan Oktay, Giacomo Tarroni, et al. “Learning Interpretable Anatomical Features Through Deep Generative Models: Application to Cardiac Remodeling”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger. Cham: Springer International Publishing, 2018, pp. 464–471 (cit. on p. 71).

[Biffi, 2020] Carlo Biffi, Juan J. Cerrolaza, Giacomo Tarroni, et al. “Explainable Anatomical Shape Analysis Through Deep Hierarchical Generative Models”. In: *IEEE Transactions on Medical Imaging* 39.6 (2020), pp. 2088–2099 (cit. on p. 95).

[Boatwright, 1980] RB Boatwright, HF Downey, FA Bashour, and GJ Crystal. “Transmural variation in autoregulation of coronary blood flow in hyperperfused canine myocardium”. In: *Circulation research* 47.4 (Oct. 1980), pp. 599–609 (cit. on p. 14).

[Bône, 2018] Alexandre Bône, Maxime Louis, Benoît Martin, and Stanley Durrleman. “Deformetrica 4: An Open-Source Software for Statistical Shape Analysis”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2018) (cit. on p. 104).

[Brugada, 2001] Josep Brugada, Luis Aguinaga, Lluìs Mont, et al. “Coronary artery revascularization in patients with sustained ventricular arrhythmias in the chronic phase of a myocardial infarction: effects on the electrophysiologic substrate and outcome”. In: *Journal of the American College of Cardiology* 37.2 (2001), pp. 529–533 (cit. on p. 96).

[BUCKBERG, 1972] GERALD D. BUCKBERG, DAVID E. FIXLER, JOSEPH P. ARCHIE, and JULIEN I.E. HOFFMAN. “Experimental Subendocardial Ischemia in Dogs with Normal Coronary Arteries”. In: *Circulation Research* 30.1 (1972), pp. 67–81. eprint: `https://www.ahajournals.org/doi/pdf/10.1161/01.RES.30.1.67` (cit. on p. 14).

[Buxton, 1999] AE Buxton, KL Lee, JD Fisher, et al. “A randomized study of the prevention of sudden death in patients with coronary artery disease. Multicenter Unsustained Tachycardia Trial Investigators”. In: *The New England journal of medicine* 341.25 (Dec. 1999), pp. 1882–1890 (cit. on pp. 71, 96).

[Cedilnik, 2018] Nicolas Cedilnik, Josselin Duchateau, Rémi Dubois, et al. “Fast personalized electrophysiological models from computed tomography images for ventricular tachycardia ablation planning”. In: *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology* 20 (Nov. 2018), pp. iii94–iii101 (cit. on pp. 53, 55).

[Cedilnik, 2020] Nicolas Cedilnik. “Image-based Personalised Models of Cardiac Electrophysiology for Ventricular Tachycardia Therapy Planning”. Theses. Université Côte d’Azur, Dec. 2020 (cit. on pp. 130, 134, 142, 145).

[Cerqueira, 2002] Manuel D. Cerqueira, Neil J. Weissman, Vasken Dilsizian, et al. “Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart”. In: *Circulation* 105.4 (2002), pp. 539–542 (cit. on p. 61).

[Chattopadhay, 2018] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N. Balasubramanian. “Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks”. In: *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018* (2018), pp. 839–847. arXiv: arXiv:1710.11063v2 (cit. on p. 71).

[Chen, 2020] Chen Chen, Chen Qin, Huaqi Qiu, et al. “Deep Learning for Cardiac Image Segmentation: A Review”. In: *Frontiers in Cardiovascular Medicine* 7 (2020) (cit. on p. 63).

[Ciaccio, 2022] Edward J. Ciaccio, Elad Anter, James Coromilas, et al. “Structure and function of the ventricular tachycardia isthmus”. In: *Heart Rhythm* 19.1 (2022), pp. 137–153 (cit. on p. 15).

[Çiçek, 2016] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. “3D U-net: Learning dense volumetric segmentation from sparse annotation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9901 LNCS (2016), pp. 424–432. arXiv: arXiv:1606.06650v1 (cit. on p. 33).

[De Tata, 1983] V De Tata, C Bergamini, Z Gori, T Locci-Cubeddu, and E Bergamini. “Transmural gradient of glycogen metabolism in the normal rat left ventricle”. In: *Pflugers Archiv : European journal of physiology* 396.1 (Jan. 1983), pp. 60–65 (cit. on p. 14).

[Dercksen, 2019] Koen Dercksen, Wouter Bulten, and Geert Litjens. “Dealing with label scarcity in computational pathology: A use case in prostate cancer classification”. In: *arXiv* (2019). arXiv: 1905.06820 (cit. on p. 71).

[Di Marco, 2017] Andrea Di Marco, Ignasi Anguera, Matthias Schmitt, et al. “Late Gadolinium Enhancement and the Risk for Ventricular Arrhythmias or Sudden Death in Dilated Cardiomyopathy: Systematic Review and Meta-Analysis”. In: *JACC: Heart Failure* 5.1 (2017), pp. 28–38 (cit. on p. 19).

[Dice, 1945] Lee R. Dice. “Measures of the Amount of Ecologic Association Between Species”. In: *Ecology* 26.3 (July 1945), pp. 297–302 (cit. on p. 37).

[Downey, 1975] J M Downey and E S Kirk. “Inhibition of coronary blood flow by a vascular waterfall mechanism.” In: *Circulation Research* 36.6 (1975), pp. 753–760. eprint: https://www.ahajournals.org/doi/pdf/10.1161/01.RES.36.6.753 (cit. on p. 14).

[Efron, 1993] Bradley Efron and Robert J. Tibshirani. “An Introduction to the Bootstrap”. In: *An Introduction to the Bootstrap*. CRC Monographs on Statistics and Applied Probability. CRC Press, 1993 (cit. on p. 79).

[Efron, 2015] Bradley Efron, David Rogosa, and Robert Tibshirani. *Resampling Methods of Estimation*. Second Edition. Vol. 19. Elsevier, 2015, pp. 492–495 (cit. on p. 81).

[Esposito, 2016] Antonio Esposito, Anna Palmisano, Sofia Antunes, et al. “Cardiac CT With Delayed Enhancement in the Characterization of Ventricular Tachycardia Structural Substrate”. In: *JACC: Cardiovascular Imaging* 9.7 (2016), pp. 822–832. eprint: https://www.jacc.org/doi/pdf/10.1016/j.jcmg.2015.10.024 (cit. on p. 19).

[Frangogiannis, 2006] Nikolaos G. Frangogiannis. “The Mechanistic Basis of Infarct Healing”. In: *Antioxidants & Redox Signaling* 8.11-12 (2006). PMID: 17034340, pp. 1907–1939. eprint: https://doi.org/10.1089/ars.2006.8.1907 (cit. on p. 14).

[Frangogiannis, 2008] Nikolaos G Frangogiannis. “The immune system and cardiac repair”. In: *Pharmacological research* 58.2 (Aug. 2008), pp. 88–111 (cit. on p. 14).

[Garreau, 2021] Damien Garreau and Dina Mardaoui. “What does LIME really see in images?” In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Proceedings of Machine Learning Research. PMLR, 2021 (cit. on p. 111).

[Gerber, 2006] Bernhard L. Gerber, Bénédicte Belge, Gabin J. Legros, et al. “Characterization of Acute and Chronic Myocardial Infarcts by Multidetector Computed Tomography”. In: *Circulation* 113.6 (2006), pp. 823–833. eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIRCULATIONAHA.104.529511 (cit. on p. 19).

[Ghannam, 2018] Michael Ghannam, Hubert Cochet, Pierre Jais, et al. “Correlation between computer tomography-derived scar topography and critical ablation sites in postinfarction ventricular tachycardia”. In: *Journal of Cardiovascular Electrophysiology* 29.3 (2018), pp. 438–445 (cit. on p. 53).

[Gilpin, 2018] Leilani H. Gilpin, David Bau, Ben Z. Yuan, et al. “Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning”. In: *CoRR* abs/1806.00069 (2018). arXiv: 1806.00069 (cit. on p. 27).

[Gong, 2018] Liyu Gong and Qiang Cheng. “Adaptive Edge Features Guided Graph Attention Networks”. In: *CoRR* abs/1809.02709 (2018). arXiv: 1809.02709 (cit. on p. 122).

[Goodfellow, 2016] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2016 (cit. on p. 22).

[Gopalan, 2022] Chaya Gopalan and Erik Kirk. “Chapter 5 - Atherosclerosis”. In: *Biology of Cardiovascular and Metabolic Diseases*. Ed. by Chaya Gopalan and Erik Kirk. Academic Press, 2022, pp. 85–101 (cit. on p. 13).

[Heidary, 2010] Shahriar Heidary, Harendra Patel, Jaehoon Chung, et al. “Quantitative Tissue Characterization of Infarct Core and Border Zone in Patients With Ischemic Cardiomyopathy by Magnetic Resonance Is Associated With Future Cardiovascular Events”. In: *Journal of the American College of Cardiology* 55.24 (2010), pp. 2762–2768 (cit. on pp. 53, 63, 96).

[Hoffman, 1985] JI Hoffman, RW Baer, FL Hanley, and LM Messina. “Regulation of transmural myocardial blood flow”. In: *Journal of biomechanical engineering* 107.1 (Feb. 1985), pp. 2–9 (cit. on p. 14).

[Hummel, 2008] Robert Hummel. “Image enhancement by histogram transformation”. In: *Computer Graphics and Image Processing* 6.2 (2008), pp. 184–195 (cit. on p. 35).

[Isensee, 2019] Fabian Isensee, Jens Petersen, Simon A. A. Kohl, Paul F. Jäger, and Klaus H. Maier-Hein. “nnU-Net: Breaking the Spell on Successful Medical Image Segmentation”. In: 1 (2019), pp. 1–8. arXiv: 1904.08128 (cit. on p. 33).

[Issa, 2018] Z. Issa, J.M. Miller, and D.P. Zipes. *Clinical Arrhythmology and Electrophysiology E-Book: A Companion to Braunwald’s Heart Disease*. Companion to Braunwald’s Heart Disease. Elsevier Health Sciences, 2018 (cit. on p. 15).

[Jia, 2018] Shuman Jia, Antoine Despinasse, Zihao Wang, et al. “Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss”. In: *Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop*. (Sept. 2018). Granada, Spain. hal-01860285v3 (cit. on pp. 33, 34, 42).

[Jiang, 2020] Xiaodong Jiang, Ronghang Zhu, Sheng Li, and Pengsheng Ji. “Co-embedding of Nodes and Edges with Graph Neural Networks”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020), pp. 1–1 (cit. on p. 122).

[Joshua, 2021] Eali Stephen Neal Joshua, Midhun Chakkravarthy, and Debnath Bhattacharyya. “Lung Cancer Detection Using Improvised Grad-Cam++ With 3D CNN Class Activation”. In: *Smart Technologies in Data Science and Communication*. Ed. by Sanjoy Kumar Saha, Paul S. Pang, and Debnath Bhattacharyya. Singapore: Springer Singapore, 2021, pp. 55–69 (cit. on p. 95).

[Jugdutt, 2003] Bodh I. Jugdutt. “Ventricular Remodeling After Infarction and the Extracellular Collagen Matrix”. In: *Circulation* 108.11 (2003), pp. 1395–1403. eprint: https://www.ahajournals.org/doi/pdf/10.1161/01.CIR.0000085658.98621.49 (cit. on p. 8).

[Jun Guo, 2020] Bang Jun Guo, Xiuxiu He, Yang Lei, et al. “Automated Left Ventricular Myocardium Segmentation Using 3D Deeply Supervised Attention U-Net for Coronary Computed Tomography Angiography; CT Myocardium Segmentation”. In: *Medical physics* 47.4 (Apr. 2020), pp. 1775–1785 (cit. on p. 53).

[Kang, 2021] Hanul Kang, Ho-min Park, Yuju Ahn, Arnout Van Messem, and Wesley De Neve. “Towards a quantitative analysis of class activation mapping for deep learning-based computer-aided diagnosis”. In: *Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*. Ed. by Frank W. Samuelson and Sian Taylor-Phillips. February 2021. SPIE, Feb. 2021, p. 20 (cit. on p. 76).

[Karim, 2021] Md. Rezaul Karim, Jiao Jiao, Till Döhmen, et al. “DeepKneeExplainer: Explainable Knee Osteoarthritis Diagnosis From Radiographs and Magnetic Resonance Imaging”. In: *IEEE Access* 9 (2021), pp. 39757–39780 (cit. on p. 95).

[Kim, 2003] Raymond J Kim, Dipan J Shah, and Robert M Judd. "How we perform delayed enhancement imaging: HOW I DO..." In: *Journal of Cardiovascular Magnetic Resonance* 5.3 (2003), pp. 505–514 (cit. on p. 19).

[Kingma, 2014] Diederik P. Kingma and Max Welling. "Auto-encoding variational bayes". In: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* Ml (2014), pp. 1–14. arXiv: 1312.6114 (cit. on p. 74).

[Kipf, 2017] Thomas N. Kipf and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks". In: *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings* (2017). arXiv: 1609.02907 (cit. on pp. 103, 106).

[Klabunde, 2011] R. Klabunde. *Cardiovascular Physiology Concepts*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 2011 (cit. on p. 16).

[Komatsu, 2013] Yuki Komatsu, Hubert Cochet, Amir Jadidi, et al. "Regional Myocardial Wall Thinning at Multidetector Computed Tomography Correlates to Arrhythmogenic Substrate in Postinfarction Ventricular Tachycardia". In: *Circulation: Arrhythmia and Electrophysiology* 6.2 (2013), pp. 342–350 (cit. on pp. 20, 53).

[Kwon, 2009] Deborah H. Kwon, Carmel M. Halley, Thomas P. Carrigan, et al. "Extent of Left Ventricular Scar Predicts Outcomes in Ischemic Cardiomyopathy Patients With Significantly Reduced Systolic Function: A Delayed Hyperenhancement Cardiac Magnetic Resonance Study". In: *JACC: Cardiovascular Imaging* 2.1 (2009), pp. 34–44 (cit. on pp. 53, 63, 96).

[Lardo, 2006] Albert C. Lardo, Marco A.S. Cordeiro, Caterina Silva, et al. "Contrast-Enhanced Multidetector Computed Tomography Viability Imaging After Myocardial Infarction". In: *Circulation* 113.3 (2006), pp. 394–404. eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIRCULATIONAHA.105.521450 (cit. on p. 19).

[Leng, 2019] Shuai Leng, Michael Bruesewitz, Shengzhen Tao, et al. "Photon-counting Detector CT: System Design and Clinical Applications of an Emerging Technology". In: *RadioGraphics* 39.3 (2019). PMID: 31059394, pp. 729–743 (cit. on p. 99).

[Lévy, 2002] Bruno Lévy, Sylvain Petitjean, Nicolas Ray, and Jérome Maillot. "Least Squares Conformal Maps for Automatic Texture Atlas Generation". In: *ACM Trans. Graph.* 21.3 (July 2002), pp. 362–371 (cit. on p. 60).

[Lin, 2020] Dongyun Lin, Yiqun Li, Shitala Prasad, et al. "CAM-UNET: Class Activation MAP Guided UNET with Feedback Refinement for Defect Segmentation". In: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, Oct. 2020, pp. 2131–2135 (cit. on p. 76).

[Lorensen, 1987] William E. Lorensen and Harvey E. Cline. "Marching Cubes: A High Resolution 3D Surface Construction Algorithm". In: *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '87. New York, NY, USA: Association for Computing Machinery, 1987, pp. 163–169 (cit. on p. 60).

[Lowekamp, 2013] Bradley C. Lowekamp, David T. Chen, Luis Ibáñez, and Daniel Blezek. "The Design of SimpleITK". In: *Frontiers in Neuroinformatics* 7 (2013) (cit. on p. 36).

[Ly, 2021] Buntheng Ly, Sonny Finsterbach, Marta Nuñez-Garcia, Hubert Cochet, and Maxime Sermesant. "Scar-Related Ventricular Arrhythmia Prediction from Imaging Using Explainable Deep Learning". In: *FIMH 2021 - 11th International Conference on Functional Imaging and Modeling of the Heart*. Vol. 12738. Lecture Notes in Computer Science. Stanford, United States: Springer International Publishing, June 2021, pp. 461–470 (cit. on p. 95).

[Maas, 2013] Andrew L. Maas, Awni Y. Hannun, and Andrew Y Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Icml '13* 28 (2013), p. 6 (cit. on p. 34).

[Mahnken, 2005] Andreas H. Mahnken, Ralf Koos, Marcus Katoh, et al. "Assessment of Myocardial Viability in Reperfused Acute Myocardial Infarction Using 16-Slice Computed Tomography in Comparison to Magnetic Resonance Imaging". In: *Journal of the American College of Cardiology* 45.12 (2005), pp. 2042–2047 (cit. on p. 20).

[McAlindon, 2015] Elisa McAlindon, Maria Pufulete, Chris Lawton, Gianni D. Angelini, and Chiara Bucciarelli-Ducci. "Quantification of infarct size and myocardium at risk: evaluation of different techniques and its implications". In: *European Heart Journal - Cardiovascular Imaging* 16.7 (Mar. 2015), pp. 738–746 (cit. on pp. 53, 96).

[Mitchell, 2003] Colleen C. Mitchell and David G. Schaeffer. "A two-current model for the dynamics of cardiac membrane". In: *Bulletin of Mathematical Biology* 65.5 (2003), pp. 767–793 (cit. on p. 142).

[Mojica, 2020] Mia Mojica, Mihaela Pop, Maxime Sermesant, and Mehran Ebrahimi. "Novel atlas of fiber directions built from ex-vivo diffusion tensor images of porcine hearts". In: *Computer Methods and Programs in Biomedicine* 187 (2020), p. 105200 (cit. on p. 136).

[Nuñez-Garcia, 2020a] M. Nuñez-Garcia, G. Bernardino, F. Alarcón, et al. "Fast Quasi-Conformal Regional Flattening of the Left Atrium". In: *IEEE Transactions on Visualization and Computer Graphics* 26.8 (2020), pp. 2591–2602 (cit. on p. 60).

[Nuñez-Garcia, 2020b] Marta Nuñez-Garcia, Nicolas Cedilnik, Shuman Jia, Maxime Sermesant, and Hubert Cochet. "Automatic multiplanar CT reformatting from trans-axial into left ventricle short-axis view". In: *STACOM 2020 - 11th International Workshop on Statistical Atlases and Computational Models of the Heart*. Lima, Peru, Oct. 2020 (cit. on p. 57).

[Nyúl, 2000] László G. Nyúl, Jayaram K. Udupa, and Xuan Zhang. "New variants of a method of MRI scale standardization". In: *IEEE Transactions on Medical Imaging* 19.2 (2000), pp. 143–150 (cit. on p. 35).

[Oktay, 2018] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, et al. "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation". In: *IEEE Transactions on Medical Imaging* 37.2 (2018), pp. 384–395 (cit. on p. 95).

[Olaf, 2015] Ronneberger Olaf, Fischer Philipp, and Brox Thomas. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* abs/1505.04597 (2015) (cit. on pp. 24, 34).

[Porter, 2009] Karen E. Porter and Neil A. Turner. “Cardiac fibroblasts: At the heart of myocardial remodeling”. In: *Pharmacology and Therapeutics* 123.2 (2009), pp. 255–278 (cit. on p. 8).

[Ribeiro, 2016] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: Association for Computing Machinery, 2016 (cit. on pp. 103, 111).

[Seeböck, 2019] Philipp Seeböck, David Romo-Bucheli, Sebastian Waldstein, et al. “Using Cyclegans for Effectively Reducing Image Variability Across OCT Devices and Improving Retinal Fluid Segmentation”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019, pp. 605–609 (cit. on p. 33).

[Simonovsky, 2017] Martin Simonovsky and Nikos Komodakis. “Dynamic edge-conditioned filters in convolutional neural networks on graphs”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3693–3702 (cit. on pp. 109, 122).

[Soejima, 2002] Kyoko Soejima, William G. Stevenson, William H. Maisel, John L. Sapp, and Laurence M. Epstein. “Electrically Unexcitable Scar Mapping Based on Pacing Threshold for Identification of the Reentry Circuit Isthmus”. In: *Circulation* 106.13 (2002), pp. 1678–1683. eprint: `https://www.ahajournals.org/doi/pdf/10.1161/01.CIR.0000030187.39852.A7` (cit. on p. 16).

[Sohn, 2015] Kihyuk Sohn, Xinchen Yan, and Honglak Lee. “Learning structured output representation using deep conditional generative models”. In: *Advances in Neural Information Processing Systems* (2015), pp. 3483–3491 (cit. on p. 71).

[Spaan, 1981] J A Spaan, N P Breuls, and J D Laird. “Diastolic-systolic coronary flow differences are caused by intramyocardial pump action in the anesthetized dog.” In: *Circulation Research* 49.3 (1981), pp. 584–593. eprint: `https://www.ahajournals.org/doi/pdf/10.1161/01.RES.49.3.584` (cit. on p. 14).

[Steenbergen, 2012] Charles Steenbergen and Nikolaos G. Frangogiannis. “Chapter 36 - Ischemic Heart Disease”. In: *Muscle*. Ed. by Joseph A. Hill and Eric N. Olson. Boston/Waltham: Academic Press, 2012, pp. 495–521 (cit. on pp. 13, 14).

[Stirrat, 2015] John Stirrat, Sebastien Xavier Joncas, Michael Salerno, Maria Drangova, and James White. “Influence of phase correction of late gadolinium enhancement images on scar signal quantification in patients with ischemic and non-ischemic cardiomyopathy”. In: *Journal of Cardiovascular Magnetic Resonance* 17.1 (Aug. 2015), p. 66 (cit. on p. 63).

[Stolzmann, 2008] Paul Stolzmann, Hans Scheffel, Sebastian Leschka, et al. “Reference values for quantitative left ventricular and left atrial measurements in cardiac computed tomography”. In: *European radiology* 18.8 (Aug. 2008), pp. 1625–1634 (cit. on p. 60).

[Sundararajan, 2017] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. “Axiomatic Attribution for Deep Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, Aug. 2017, pp. 3319–3328 (cit. on pp. 103, 111, 113).

[Suzuki, 2017] Kenji Suzuki. “Overview of deep learning in medical imaging”. In: *Radiological Physics and Technology* 10.3 (Sept. 2017), pp. 257–273 (cit. on p. 26).

[Takigawa, 2019] Masateru Takigawa, Josselin Duchateau, Frederic Sacher, et al. “Are wall thickness channels defined by computed tomography predictive of isthmuses of postinfarction ventricular tachycardia?” In: *Heart Rhythm* 16.11 (2019), pp. 1661–1668 (cit. on pp. 53, 64, 97).

[Thompson, 1942] Darcy Wentworth Thompson and D’Arcy W Thompson. *On growth and form*. Vol. 2. Cambridge university press Cambridge, 1942 (cit. on p. 104).

[Unal, 2017] Ilker Unal. “Defining an optimal cut-point value in ROC analysis: An alternative approach”. In: *Computational and Mathematical Methods in Medicine* 2017 (2017) (cit. on p. 84).

[Valette, 2008] Sebastien Valette, Jean Marc Chassery, and Remy Prost. “Generic Remeshing of 3D Triangular Meshes with Metric-Dependent Discrete Voronoi Diagrams”. In: *IEEE Transactions on Visualization and Computer Graphics* 14.2 (2008), pp. 369–381 (cit. on p. 60).

[Vedaldi, 2008] Andrea Vedaldi and Stefano Soatto. “Quick shift and kernel methods for mode seeking”. In: *In European Conference on Computer Vision, volume IV*. 2008, pp. 705–718 (cit. on p. 111).

[Xia, 2005] Yunlong Xia, Ole Kongstad, Eva Hertervig, et al. “Activation recovery time measurements in evaluation of global sequence and dispersion of ventricular repolarization”. In: *Journal of electrocardiology* 38.1 (Jan. 2005), pp. 28–35 (cit. on p. 145).

[Yaniv, 2018] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research”. In: *Journal of Digital Imaging* 31.3 (June 2018), pp. 290–303 (cit. on p. 36).

[Yezzi, 2003] Anthony J. Yezzi and Jerry L. Prince. “An Eulerian PDE Approach for Computing Tissue Thickness”. In: *IEEE Transactions on Medical Imaging* 22.10 (2003), pp. 1332–1339 (cit. on p. 58).

[Zegard, 2021] Abbasin Zegard, Osita Okafor, Joseph de Bono, et al. “Myocardial Fibrosis as a Predictor of Sudden Death in Patients With Coronary Artery Disease”. In: *Journal of the American College of Cardiology* 77.1 (2021), pp. 29–41 (cit. on pp. 53, 63, 96).

[Zhang, 2021] Xiao Meng Zhang, Li Liang, Lin Liu, and Ming Jing Tang. “Graph Neural Networks and Their Current Applications in Bioinformatics”. In: *Frontiers in Genetics* (2021) (cit. on p. 103).

[Zhou, 2016] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. “Learning Deep Features for Discriminative Localization”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2004. 1. IEEE, June 2016, pp. 2921–2929 (cit. on p. 76).

[Zhu, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2242–2251 (cit. on p. 33).