Cardiovasculaire medische beeldanalyse met 'Deep Learning'

Cardiovascular Medical Image Analysis with Deep Learning

Patricio Astudillo

Promotoren: prof. dr. ir. J. Dambre, dr. ir. P. Mortier
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: biomedische ingenieurstechnieken

UNIVERSITEIT
GENT

Vakgroep Elektronica en Informatiesystemen
Voorzitter: prof. dr. ir. K. De Bosschere
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2019 - 2020

ISBN 978-94-6355-364-3
NUR 984, 870
Wettelijk depot: D/2020/10.500/41

**Members of the exam committee**

| | |
|---|---|
| Ereprof dr. ir. | Ronny Verhoeven, *chairman*<br>Faculty of Engineering and Architecture<br>Ghent University |
| prof. dr. ir. | Patrick Segers<br>Faculty of Engineering and Architecture<br>Ghent University |
| prof. dr. ir. | Francis wyffels<br>Faculty of Engineering and Architecture<br>Ghent University |
| prof. dr. | Roel Van Holen<br>Faculty of Engineering and Architecture<br>Ghent University |
| prof. dr. | Jan D'hooge<br>Department of Cardiovascular Sciences<br>KU Leuven |
| prof. dr. | Peter de Jaegere<br>Department of Cardiology<br>Erasmus Medical Center, Rotterdam |
| dr. ir. | Matthieu De Beule<br>FEops, Ghent |

# Acknowledgements

First of all, I would like to thank my promotor Joni Dambre, whose sobriety and honesty towards my work kept me focused. I am also grateful that she accepted to take over my PhD from my previous promotor.

Next, I would like to thank all people from FEops and especially the following people. First of all Matthieu and Peter, who allowed me to pursue a PhD at FEops within the scope of a Marie Curie scholarship. Special thanks to Peter, whose advice and proof-reading improved my work significantly and who always made time for me. Next, I would like to thank professor Verhegghe, who helped me receiving the necessary GPUs through the grant program of NVIDIA. If it was not for the following people, I could not have done my research: Alessandra, Elisa, Giorgia, Mariana, Nieves, Thijs and Tim. Thanks for analysing all the images of the patients and providing me with the necessary (expert) data and insights. A special thanks go out to Francesco, who traded the proofreading of my first draft for a lesson in diplomacy at a European meeting in Greece. A special thanks to Nic, who inspired me to develop my photoshop skills.

Next, I would like to thank the following people with whom I've had the opportunity and pleasure to collaborate with. First of all, I would like to thank the co-authors dr. Ole De Backer, prof. dr. Johan Bosmans and prof. dr. Peter de Jaegere. Their clinical insights improved my work, and they provided the necessary image and clinical data for some of my studies. Next, I would like to thank dr. Cameron Dowling for his inspiring enthusiasm in science and providing me with the necessary data and insights for the classification study. Thirdly, I would like to express my gratitude to dr. Anish Bhuva for allowing me to perform my secondment at the University College of London. Next, I would like to thank Prof. Francis wyffels for his assistance during the development of the DQL study and with whom I enjoyed working with. Finally, special thanks to Jonas De Grave for his initial guidance and input during the optimisation of my first deep learning models and the hackathon experience.

Special thanks to the people from the writing group, especially Rosa, Melissa, Celine and Kaat. The writing weeks pushed my work forward and improved my snow-man-building-skills significantly. I hope my coffee machine made up for my "focussed" mood.

Since I'm only standing on the shoulders of giants, I would like to thank the following people in my prior education: Kristel Balcaen, Kevin Hoefman, Peter Verswyvelen, prof. dr. Elisa Boix and dr. Carlos Noguera. Special thanks to prof. dr. Wolfgang De Meuter, prof. dr. Coen De Roover and prof. dr. em. Dirk Vermeir who provided the necessary recommendations.

I want to express my gratitude to all my friends, who supported me in their original way. Bart for his warm friendship and his audaciousness scientific standpoints. Kathy for introducing ashtanga to me and reminding me that lake-swimming is not only a summer activity. Elena and Lisa, whose marriage still inspires me to appreciate the important things in life. Joppe and his family who gave me a god-child Flore. Josie for her kind but clear support and the necessary escapes to Nice or Barcelona.

Morre who I only saw on the snowboard tracks. Special thanks to Muroen, Wouter, Sofie, Milan, Jannis, Jesse, Hannah, Niko, Roel, Maya and Lilith whose friendship and humour kept me on my feet (both ways).

Tenslotte wil ik mijn familie bedanken, zonder hen zou ik nu niet staan waar ik vandaag sta. Mijn moeder voor haar ondersteuning bij zowat elke grote stap in mijn leven. Merci ma! Mijn vader, wiens hard verdiende positiviteit me inspireerde om door te gaan tijdens de moeilijkere momenten van de afgelopen vier jaar. Mijn grootmoeder en mijn grootste inspiratiebron om de wereld van de wetenschap in te duiken en die mijn diplomatische kant gaandeweg bijvijlde met geduld en dankbaarheid. Mijn grote broer en held wiens advies en ondersteuning de doorslag gaven om mijn leven aan de Côte d'Azur op te geven.

*Ghent, May 2020,*
*Patricio Astudillo*

# Samenvatting

Dit doctoraatsproefschrift gebruikt artificiële intelligentie (AI) om cardiovasculaire medische beelden te analyseren en beoogt daarbij efficiënte en nauwkeurige modellen aan te bieden die artsen kunnen helpen bij het besluitvormingsproces. De belangrijkste focus van deze studie ligt op transkatheter hartklepinterventies, een behandeling waarvan verwacht wordt dat ze de komende jaren aanzienlijk zal toenemen. Omdat deze ingrepen gewoonlijk via de lies gebeuren en de borstkas hiervoor niet geopend moet worden, is een nauwkeurige preoperatieve medische beeldanalyse cruciaal voor een optimale behandeling van de patiënt. De analyse is echter niet altijd eenvoudig vanwege de complexiteit van specifieke pathologieën, dimensies van de beelden, beeldkwaliteit en de ervaring van de operator. Voor experten zijn bicuspide aortakleppen bijvoorbeeld nog steeds een pathologie die moeilijk te detecteren is op een driedimensionaal (3D) medisch beeld. Automatische beeldanalyse kan hier mogelijks een oplossing bieden. Bovendien kan automatisatie helpen om de efficiëntie te verhogen in hoge volume centra met veel ervaring, en kan het helpen om bij minder ervaren centra de nauwkeurigheid van de analyses te verhogen.

Om tegemoet te komen aan deze noden, heb ik verschillende van deze transkatheter hartklepinterventies onderzocht en automatische technieken bedacht, gebaseerd op deep learning, die de arts kunnen helpen met cardiovasculaire medische beeldanalyse.

### Morfologie classificatie van de aortaklep

In deze studie stelden we dat deep learning kan gebruikt worden om tricuspide en bicuspide aortakleppen (BAV) van patiënten te classificeren uit MDCT beelden. BAV is de meest voorkomende aangeboren hartaandoening bij volwassenen. Bovendien is de aanwezigheid van BAV een belangrijke factor bij het selecteren van de juiste behandeling voor patiënten met symptomatische ernstige aortaklepstenose. Experten beweren dat het moeilijk is om deze pathologie te classificeren omdat de verkalking van de klepblaadjes de juiste classificatie kunnen verhinderen en de beeldvormende modaliteit mogelijks niet de details van de pathologie onthult.

Deze retrospectieve studie werd uitgevoerd met MDCT data van patiënten die waren doorverwezen voor minimaal invasieve aortaklepvervanging (TAVR). Een 3D convolutionair neuraal netwerk (CNN) werd ontwikkeld om de MDCT afbeeldingen te classificeren als tricuspide of bicuspide. De cross-validatie van de dataset toonde aan dat een getraind 3D CNN model een uitstekend vermogen had om de aortaklep correct te classificeren, met een hoge sensitiviteit (100 %) en specificiteit (92 %). Deze resultaten waren zeer gunstig, aangezien het model op een relatief kleine dataset werd getraind. Gezien deze optimale resultaten, kunnen we stellen dat een dergelijk hulpmiddel door clinici kan gebruikt worden om BAV te identificeren. Als we naar de beschikbare literatuur kijken, zijn er geen studies die een 3D CNN model hebben gebruikt om een volledig 3D volume als input te classificeren.

### Geautomatiseerde selectie van de TAVR-implantaat grootte

Verschillende metingen van de aorta annulus zijn noodzakelijk tijdens het besluitvormingsproces dat voorafgaat aan een TAVR-procedure. De uiteindelijke afmeting van de gekozen prothese wordt mede bepaald door de oppervlakte of de omtrek van deze aorta annulus. Deze ring wordt gemeten op het annulaire vlak van de aorta (AAP); een 2D vlak gedefinieerd door de onderste scharnierpunten van de aortaklepbladen.

In deze retrospectieve studie werden de MDCT data van 355 patiënten gebruikt om twee CNN-modellen te trainen om de aorta annulus te segmenteren. De oppervlakte en de omtrek werden afgeleid van de voorspelde segmentatie en werden gebruikt om de implantaat grootte te selecteren van de Edwards Sapien 3 en de Medtronic Evolut kleppen. De validatie werd in drie stappen uitgevoerd. Eerst werd een test-dataset van 118 patiënten gebruikt om de nauwkeurigheid te evalueren. Deze patiënten werden uitgesloten van de trainingsfase. De resultaten toonden aan dat de oppervlakte en de omtrek op een automatische, reproduceerbare, efficiënte en nauwkeurige manier konden worden voorspeld aan de hand van onze methode. Ten tweede waren de verschillen tussen de handmatig verkregen metingen van de aorta annulus en de automatische voorspellingen vergelijkbaar met de verschillen tussen twee onafhankelijke operatoren. Dit duidt op een voldoende nauwkeurigheid van onze voorgestelde aanpak. Ten slotte werd bij het invoegen van de voorspelde oppervlakte en omtrek in de maattabellen van de Edwards Sapien 3 en het Medtronic Evolut kleppen, gevonden dat de automatisch geselecteerde klepmaten goed overeenkwamen met de handmatig bepaalde klepmaten. De totale analysetijd van AAP tot de grootte van de prothese was minder dan 1 seconde, wat veel sneller is dan de handmatige verwerkingstijd van ongeveer 10 minuten. Daarom kan onze voorgestelde methode de huidige pre-operatieve TAVR planning versnellen en analysefouten in nieuwe centra verminderen.

### Op weg naar een veilige en efficiënte preoperatieve planning van minimale invasieve mitralisklepvervanging (TMVR)

De detectie van de annulaire ring van de mitralisklep is een tijdrovend en uitdagend proces. De mitralisklep-annulus heeft een zadelvormige anatomie, die in grootte en vorm verandert tijdens de hartcyclus. Van de gedetecteerde mitralisklep-annulus kunnen aanvullende metingen worden bepaald, zoals de oppervlakte, 2D en 3D perimeter, trigone-naar-trigone-, septaal-lateraal- en commissure-naar-commissure-afstand. De annulaire ring van de mitralisklep en de afgeleide metingen daarvan zijn cruciaal tijdens de TMVR-procedure en ontwikkeling van nieuwe mitralisklepprotheses.

Deze retrospectieve studie gebruikte de MDCT beelddata en expertgegevens van 71 patiënten om deep learning modellen te trainen om de mitralisklep-annulus te segmenteren uit de medische beelden. Eenmaal gesegmenteerd, werden de bovenstaande metingen afgeleid van de mitralisklep-annulus tijdens een post-processing stap. De methode werd cross-gevalideerd vanwege de kleine omvang van de dataset. De nauwkeurigheid van de methode werd bevestigd door de voorspelde en handmatig afgeleide metingen te vergelijken. De totale analysetijd van onze voorgestelde methode was minder dan 1 seconde per patiënt, terwijl de handmatige verwerkingstijd rond de 25 minuten per patiënt ligt. De efficiëntie en nauwkeurigheid van onze voorgestelde methode bieden het vertrouwen deze technologie naar de klinische praktijk te brengen. Een voorstel voor een klinische toepassing voor medische beeldanalyse werd gepresenteerd in deze studie.

### Detectie van het annulaire vlak van de aortaklep en de coronaire ostia.

Tijdens de preoperatieve planning van een TAVR-procedure, worden de MDCT-afbeeldingen gebruikt om bepaalde anatomische risicofactoren te identificeren. Zo worden bijvoorbeeld de afstanden van de AAP tot de linker en rechter coronaire ostium (LCO, RCO) geïdentificeerd. Deze afstanden helpen om het risico op coronaire obstructie te begrijpen. Deze obstructie is een potentieel levensbedreigende complicatie waarbij de bloedstroom naar een kransslagader aanzienlijk wordt verminderd. De AAP wordt gedefinieerd door de drie onderste scharnierpunten van de aortaklepbladen: de linker-, non- en rechter coronaire cusp (LCC, NCC en RCC).

In deze retrospectieve studie werden de MDCT beelddata en expertgegevens van 344 patiënten gebruikt om modellen te trainen die de vijf bovengenoemde oriëntatiepunten konden detecteren. De methode combineert de resultaten van drie getrainde CNN-modellen en een post-processing stap. Drie stappen werden toegepast om de methode te valideren. Eerst werd een test-dataset van 100 patiënten (deze werden uitgesloten van de trainingsfase) gebruikt om de nauwkeurigheid van de methode te evalueren. De resultaten toonden aan dat de vijf oriëntatiepunten efficiënt en nauwkeurig konden worden gedetecteerd. Vervolgens werd een interobserver-variabiliteitsstudie uitgevoerd. Het verschil tussen de handmatig en automatisch geïdentificeerde oriëntatiepunten was over het algemeen kleiner in vergelijking met verschillen die werden waargenomen tussen twee onafhankelijke operatoren. Dit geeft aan dat de voorgestelde aanpak deze oriëntatiepunten met acceptabele nauwkeurigheid detecteert. Ten slotte werd ook geïllustreerd dat de methode het mogelijk maakt om klinisch relevante metingen zoals coronaire hoogte nauwkeurig en automatisch te bepalen. De totale analysetijd van MDCT-afbeelding naar de voorspelde punten was minder dan 1 seconde, wat duidelijk het potentieel van onze voorgestelde methode aantoont om de huidige pre-operatieve planning te versnellen.

### Curriculum deep learning met verschillende exploratiestrategieën: een haalbaarheidsstudie naar de detectie van cardiovasculaire oriëntatiepunten.

In deze retrospectieve studie werd een curriculum deep q-learning algoritme gebruikt om de inferieure rand van het membraneuze septum automatisch te detecteren. Deze rand is een bekende anatomische markering voor het atrioventriculaire geleidingspad en werd in onze studie gedefinieerd door drie punten. De preoperatieve beoordeling van de locatie van het atrioventriculaire geleidingssysteem is belangrijk om het risico op geleidingsstoornissen na een TAVR-procedure beter te begrijpen en zelfs te voorspellen.

Het idee achter curriculum deep q-learning is dat een artificiële agent geleidelijk de oriëntatiepunten in het medische beeld leert te vinden. De agent heeft een beperkt zicht en leert naar de oriëntatiepunten te lopen. Tijdens de wandeling probeert het de notie van cumulatieve beloning te maximaliseren. Het model werd cross-gevalideerd vanwege de kleine dataset (278 patiënten). De validatie illustreerde dat de methode nauwkeurig en efficiënt was. Een interobserver-variabiliteitsstudie bevestigde de nauwkeurigheid van de methode. De output van dit onderzoek was drieledig: ten eerste werden de voorspelde locaties nauwkeurig en efficiënt gedetecteerd. Vervolgens werden twee nieuwe strategieën voor actie-selectie $\alpha$-decay en actie-dropout gevalideerd en vergeleken met de reeds bestaande $\epsilon$-decay. Ten slotte bleek de taak van het detecteren van deze oriëntatiepunten computationeel goedkoop (in vergelijking met vergelijkbare methoden), maar moeilijk genoeg om als basis te dienen om te

experimenteren met nieuwe DQL technieken.

# Summary

This doctoral thesis involves AI to analyse cardiovascular medical images and aims to provide efficient and accurate models that can assist the physician in the decision-making process. The main focus lies with transcatheter heart valve interventions where the number of treatments and patients is expected to increase significantly in the coming years. Since these interventions are closed-chest procedures, accurate pre-operative medical image analysis is crucial for optimal patient treatment. However, the analysis is not always straightforward due to the complexity of specific pathologies, image dimensions, image quality and the operator experience. For example, bicuspid aortic valves are still, for the experts' eye, a pathology that is challenging to detect on a three-dimensional (3D) image. Automatic image analysis may offer a solution here. Moreover, automation can help to increase efficiency in experienced high-volume centers, and this can also help in less experienced centers to increase the accuracy of the analyses.

To meet these needs, I have investigated several of these transcatheter heart valve interventions and devised automatic techniques, based on deep learning, that can assist the physician with cardiovascular medical image analysis.

### Aortic valve morphology classification

In this study, we hypothesised that deep learning could be used to classify tricuspid and bicuspid aortic valve (BAV) patients from MDCT images. BAV is the most common congenital heart disorder in adults. The presence of BAV is a crucial factor when selecting the proper treatment for patients with symptomatic severe aortic stenosis who are at increased risk for surgery. Experts claim that it is difficult to classify this pathology, as the leaflet calcification may prevent proper classification and the imaging modality may not reveal the details of the pathology.

This retrospective study was performed with the MDCT data of patients who had been referred for transcatheter aortic valve replacement (TAVR). A three-dimensional (3D) convolutional neural network (CNN) was developed to classify the MDCT images and label the aortic valve morphology as tricuspid or bicuspid. The cross-validation of the dataset showed that a trained 3D CNN had an excellent ability to classify aortic valve morphology, with a high degree of both sensitivity (100%) and specificity (92%). These results are very favourable, especially given the relatively small dataset on which the model was trained. Given these features, it is foreseeable that a tool such a this could be used by clinicians to identify BAV. When looking at the available literature, there are no studies which have used a 3D CNN model to classify entire 3D region of interest (ROI) as input.

### Enabling automated device size selection for TAVR

Different measurements of the aortic annulus are crucial during the decision-making process prior to a TAVR procedure. For example, the area or perimeter of the

measured aortic annulus determine (among other measurements) the final prosthesis device size. The aortic annulus is measured on the aortic annular plane (AAP); a two-dimensional (2D) plane defined by the lowest hinge points of the aortic valve's leaflets.

In this retrospective study, the MDCT images of 355 patients were used to train two CNN models to segment the aortic annulus. The area and perimeter were derived from the predicted segmentation and were used to select prosthesis sizes for the Edwards Sapien 3 and the Medtronic Evolute device. The validation was performed in three steps. First, a test dataset of 118 patients was used to evaluate accuracy. These patients were excluded from the training phase of the models. Results showed that the area and perimeter could be predicted in an automatic, reproducible, efficient and accurate manner using our method. Secondly, the differences between the manually obtained aortic annulus measurements and the automatic predictions were similar to the differences between two independent observers. This indicates a satisfying accuracy of the proposed approach. Finally, when inserting the predicted area and perimeters in the sizing charts of the Edwards Sapien 3 and the Medtronic Evolute device, it was found that the automatically selected device sizes accorded well with the manually determined device sizes. The total analysis time from AAP to prosthesis size was below 1 second, which is much faster than the manual processing time of approximately 10 minutes. Therefore, our proposed method can speed-up the current pre-operative TAVR planning as well as reduce analysis errors in new centers.

### Towards safe and efficient pre-operative planning of transcatheter mitral valve replacement (TMVR)

The detection of the mitral valve annulus is a time-consuming and challenging process. The mitral valve annulus has a saddle-shaped anatomy, which changes in size and shape during the cardiac cycle. From the detected mitral valve annulus, additional measurements can be assessed such as the area, 2D and 3D perimeter, trigone-to-trigone-, septal-lateral- and commissure-to-commissure distance. The mitral valve annulus and the derived measurements thereof are crucial during the TMVR procedure and mitral valve prosthesis development.

This retrospective study used the MDCT images and expert data of 71 patients to train deep learning models to segment the mitral valve annulus from medical images. Once segmented, the measurements above were derived from the mitral valve annulus during a post-processing step. The method was cross-validated due to the small size of the dataset. The accuracy of the method was confirmed by comparing the predicted and manually derived measurements. The total analysis time of our proposed method was less than 1 second per patient, which is much faster the manual processing time of approximately 25 minutes per patient. The efficiency and accuracy of our proposed method provide the confidence to move towards implementation of this technology in clinical practice. A proposal for a clinical application for medical image analysis was presented in this study.

### Aortic annular plane and coronary ostia detection

During the pre-operative planning of a TAVR procedure, the MDCT images are used to identify certain anatomical risk factors. For example, the distances from the AAP to the left coronary ostium (LCO) and the right coronary ostium (RCO) are identified. These distances help to understand the risk of coronary obstruction. This obstruction is a potentially life-threatening complication during which the blood flow to a coronary artery is reduced significantly. The AAP is defined by the three

lowest hinge points of the aortic valve's leaflets: the left coronary cusp (LCC), the non coronary cusp (NCC) and the right coronary cusp (RCC).

In this retrospective study, the MDCT images and expert data of 344 patients were used to train models that could detect the five landmarks mentioned above. The method combined the results of three trained CNN models and a post-processing step. Three steps were applied to validate the method. First, a test dataset of 100 patients (these were excluded from the training phase) was used to evaluate the accuracy of our method and results showed that the five landmarks could be detected efficiently and accurately. Next, an inter-observer variability study was conducted. The difference between the manually and automatically identified landmarks was generally smaller when compared to differences observed between two independent operators. This indicates that the suggested approach detects these landmarks within acceptable accuracy. Finally, it was also illustrated that the method allows to accurately and automatically determine clinically relevant measurements such as coronary height. The total analysis time from MDCT image to the predicted landmarks was less than one second, which clearly shows the potential of our proposed method to speed up current pre-operative planning workflows.

### Curriculum deep reinforcement learning with different exploration strategies: A feasibility study on cardiac landmark detection

In this retrospective study, a curriculum deep q-learning (DQL) algorithm was used to detect the inferior border of the membranous septum automatically. This border is a known anatomical marker for the atrioventricular conduction path and was defined in our study by three points. The pre-operative assessment of the location of the atrioventricular conduction path is crucial to understand the risk of conduction abnormalities following a TAVR.

The idea behind curriculum DQL is that an artificial agent gradually learns to find the landmarks within the medical image. The agent has a limited view and learns to walk towards the target landmarks. During the walk, it tries to maximise some notion of cumulative reward. The model was cross-validated due to the small size of the dataset consisting of 278 patients. The validation illustrated that the method was accurate and efficient. An inter-observer variability study confirmed the accuracy of the method. The output of this study was three-fold: first, the predicted landmark locations were detected accurately and efficiently. Next, two novel action-selection strategies $\alpha$-decay and action-dropout were validated and compared to the already established $\epsilon$-decay. Finally, the task of detecting these landmarks was computationally cheap (when compared to similar methods) yet difficult enough to serve as a baseline to experiment with novel DQL techniques.

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| 1D | one-dimensional. |
| 2D | two-dimensional. |
| 3D | three-dimensional. |
| 4D | four-dimensional. |
| AAP | aortic annular plane. |
| Adam | adaptive moment estimation. |
| AI | artificial intelligence. |
| AS | aortic stenosis. |
| AUC | area under the roc curve. |
| AV | atrioventricular. |
| BAV | bicuspid aortic valve. |
| CAC | coronary artery calcification. |
| CAGR | compound annual growth rate. |
| CNN | convolutional neural network. |
| CoA | coarctation of the aorta. |
| CS | cardiac sarcoidosis. |
| CT | computed tomography. |
| DBN | deep belief network. |
| DCGAN | deep convolutional generative adversarial networks. |
| DL | deep learning. |
| DNN | deep neural network. |
| DQL | deep q-learning. |
| DQN | deep q-network. |
| ECG | electrocardiography. |
| EDV | end-diastolic volume. |
| EF | ejection fraction. |
| ESV | end-systolic volume. |
| FCNN | fully convolutional neural network. |
| FN | false negative. |
| FP | false positive. |
| GAN | generative adversarial network. |
| GPU | graphical processing unit. |
| HU | Hounsfield units. |

| | |
|---|---|
| LA | left atrium. |
| LCC | left coronary cusp. |
| LCO | left coronary ostium. |
| Leaky ReLU | leaky rectified linear unit. |
| LV | left ventricle. |
| LVH | left ventricle hypertrophy. |
| LVOT | left ventricular outflow tract. |
| MAC | mitral annular calcification. |
| MDCT | multidetector computed tomography. |
| ML | machine learning. |
| MR | magnetic resonance. |
| MSE | mean squared error. |
| NCC | non coronary cusp. |
| NLL | negative log-likelihood. |
| OCT | optical coherence tomography. |
| PET | positron emission tomography. |
| RA | right atrium. |
| RCC | right coronary cusp. |
| RCO | right coronary ostium. |
| ReLU | rectified linear unit. |
| RI | retinal images. |
| RNN | recurrent neural network. |
| ROI | region of interest. |
| ROM | random orthogonal matrices. |
| RV | right ventricle. |
| SAVR | surgical aortic valve replacement. |
| SGD | stochastic gradient descent. |
| STS-PROM | society of thoracic surgeons predicted risk of mortality. |
| SV | stroke volume. |
| TAVR | transcatheter aortic valve replacement. |
| TMVR | transcatheter mitral valve replacement. |
| TN | true negative. |
| TP | true positive. |
| US | ultrasound. |
| WHO | world health organization. |

# 1 Introduction

*Artificial intelligence*
*is the new electricity*

Andrew Ng

Artificial intelligence (AI) has known an unprecedented rise in the past decade, including medical image analysis. This extraordinary progress was made possible through substantial advancements in its subfields: machine learning (ML) and deep learning (DL). Current algorithmic performances surpass similar algorithms or even humans by a large margin, and although many are enthusiastic about this scientific achievement, others fear what the future might bring. AI has already known uprises and downfalls [1] in the past, and therefore, it is essential to temper the current enthusiasm by matching it with realism, including the domain of medical image analysis. AI can assist the human physician with the medical image analysis by providing efficient and accurate models that encapsulate expert knowledge. However, the human aspect and human decision making in health care will always be crucial [1, 2].

This doctoral thesis involves AI to analyse cardiovascular medical images and aims to provide efficient and accurate models that can assist the physician in the decision-making process. The main focus lies with transcatheter heart valve interventions where the number of treatments and patients is expected to increase significantly in the coming years. Since these interventions are closed-chest procedures, accurate pre-operative medical image analysis is crucial for optimal patient treatment. However, the analysis is not always straightforward due to the complexity of specific pathologies, image dimensions, image quality and the operator experience. For example, bicuspid

[1] The so-called 'AI-winters'

aortic valves are still, for the experts' eye, a pathology that is challenging to detect on a three-dimensional (3D) image. Automatic image analysis may offer a solution here. Moreover, automation can help to increase efficiency in experienced high-volume centers, and this can also help in less experienced centers to increase the accuracy of the analyses.

To meet these needs, I have investigated several of these transcatheter heart valve interventions and devised automatic techniques, based on deep learning, that can assist the physician with cardiovascular medical image analysis.

## 1.1 FEops

This PhD was supported by the European Horizon 2020 Framework[2] through the MSCA-ITN[3] and was part of the MUSICARE[4] PhD training program. The main goal of MUSICARE was to create a bridge between academia, industry and physicians in order to foster research and develop technology in the cardiovascular domain. One of the partners of this project was FEops, which is where I conducted this PhD research.

FEops started as a spin-off company from UGent and is now a company that provides personalised computational modelling and simulation for structural heart interventions. FEops HEARTguide™ is a platform that uses advanced simulation technology to provide physicians and device manufacturers pre-operative insights into the interaction between a device and the specific patient anatomy. These insights may accelerate research and development of novel transcatheter-based solutions and have the potential to improve clinical outcomes in a hospital setting.

FEops HEARTguide™ encapsulates the following workflow: first, a physician uploads medical images of a patient. Upon receiving these images, trained case analysts inspect the images by rating the image quality and perform pre-processing steps such as cardiac phase selection, region of interest selection and anatomical measurements. This step is crucial for the subsequent steps and the final report. Next, the simulations are carried out. The interaction of the device within the anatomy is simulated during the finite element analysis (FEA) and, in case of aortic heart valve replacement, computed fluid dynamics (CFD) simulation is used to predict potential valve leakage. Finally, a report summarises the image analysis and the output of the simulations and is made available online where the physician can inspect the report. An overview of this workflow is depicted in Figure 1.1.

[2]European Horizon 2020 Framework project number 642458
[3]Marie Sklodowska-Curie Action International Training Network
[4]MUltiSectorial Integrative approaches to CArdiac caRE: www.musicare2020.eu

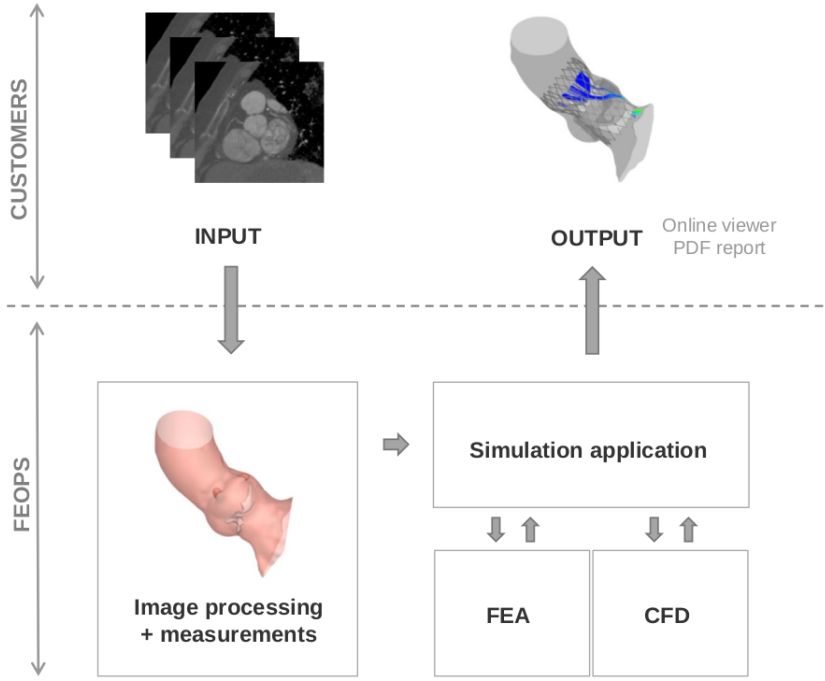



*Figure 1.1: Overview of the FEops HEARTguide$^{TM}$ workflow. The workflow starts when a customer uploads medical images. Next, image processing and anatomical measurements are performed by trained case analysts. After the simulations, a report is sent back to the customer.*

The first patient-specific modelling workflow of FEops focused on the interaction between a prosthesis and the aortic valve anatomy during a transcatheter aortic valve replacement (TAVR). During the past years, FEops has performed the above-described workflow to a significant number of patients to simulate a TAVR procedure. The medical images of these patients and corresponding expert annotations were the starting point of my research in the domain of the aortic valve. In this thesis, we introduce novel methods that automate a few of the TAVR pre-operative image analyses.

## 1.2 Aortic valve

The aortic valve lies between the left ventricle and the aorta. This semilunar heart valve normally has three leaflets which prevent the blood from flowing backwards during ventricular diastole. Aortic stenosis (AS) is a progressive valvular heart disease that reduces the motion of the aortic leaflets and valve area [3]. The conventional treatment for aortic stenosis is open-chest surgical aortic valve replacement (SAVR) but for patients at high- and intermediate surgical risk, a minimally invasive alternative,

called TAVR, has become a well-accepted procedure [4–6].

TAVR is a closed-chest procedure, and in order to assess the anatomical properties, a heart team analyses the diseased aortic valve pre-operatively with CT images [7]. During a TAVR, a crimped prosthetic valve is positioned in the aortic root and deployed. The expanding metallic frame crushes the calcified native leaflets against the aortic wall and, prosthetic leaflets attached to this frame take over the valve's function (Figure 1.2).

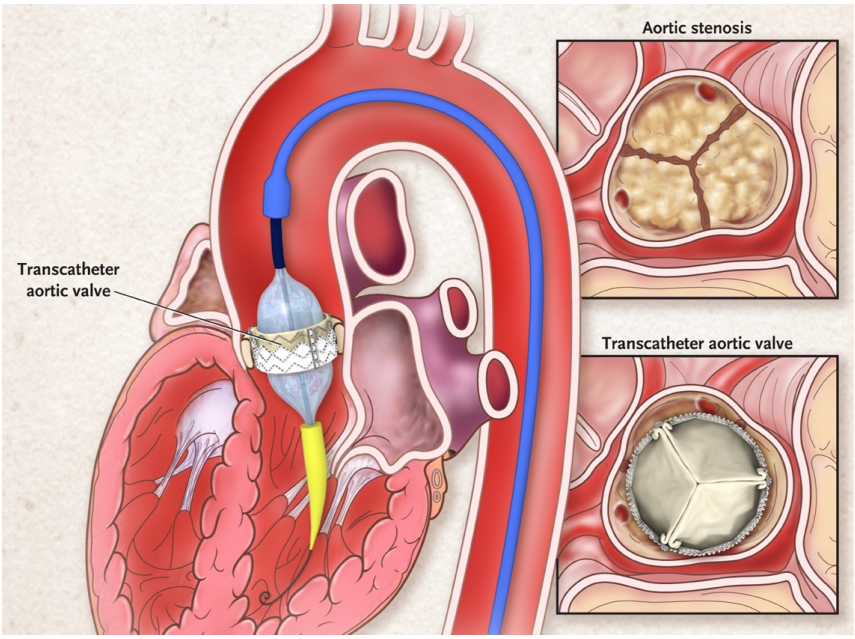



*Figure 1.2: Schematic overview of TAVR. The transcatheter aortic valve is positioned in the aortic root and deployed. Image from [4].*

TAVR has known a significant evolution in the past decade, evolving from a last resort procedure for an inoperable patient [8], to a proper alternative to the surgical approach for high-risk patients [9–11]. Soon thereafter, studies concluded that treating intermediate-risk patients using TAVR has a similar outcome as the surgical approach [12, 13], and recent studies showed that TAVR is at least as good as the surgical approach in low-risk patients [14, 15]. Including high-, intermediate- and low-risk patients as candidates for TAVR will significantly increase the number of patients undergoing this treatment. This evolution is reflected in the expected global TAVR market revenue which is expected to grow towards $8 billion by 2025[5].

[5]https://www.alliedmarketresearch.com/tavi-market

## 1.3 Mitral valve

Efforts have been made to apply the knowledge of transcatheter aortic valve solutions to the field of other heart valves such as the mitral valve. Mitral valve regurgitation is the most common valvular disease in the western world, affecting 1 in 10 individuals aged 75 and older [16, 17]. Regurgitation is the inability of a valve to close completely, which influences the prevention of the back-flow of the blood. Currently, the standard procedure for mitral valve regurgitation is the surgical approach. However, similar to the story of TAVR, high-risk patients are excluded from this invasive surgery. Percutaneous mitral valve repair can be a viable alternative for this high-risk patient group [18], but recent data also shows that transcatheter mitral valve replacement (TMVR) is an emerging option [17, 19].

The complexity of the mitral valve is one of the reasons that transcatheter-based solutions in this field are moving slower. The mitral valve has a non-planar, saddle-shaped annulus which deforms significantly during the cardiac cycle. This anatomical shape is difficult to determine, but the adequate characterisation is crucial for research and clinical purposes. Figure 1.3 depicts a simplified image of the mitral valve. Since it is important to capture the movement of the mitral valve annulus, volumetric CT images are used to capture the cardiac cycle resulting in four-dimensional (4D) (3D + time) medical images. In this thesis, we discuss a novel method to detect the mitral valve annulus from medical images automatically.

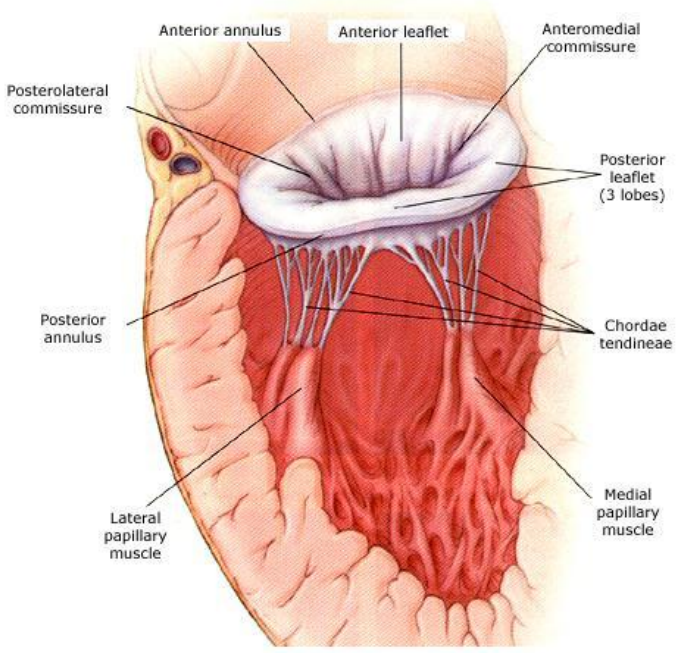



*Figure 1.3: The human heart with a focus on the mitral valve apparatus. Image from www.heart-valve-surgery.com.*

## 1.4 Medical image analysis

The aortic- and mitral transcatheter-solutions require medical image analysis in order to understand the anatomy and pathology of the valve. In the domain of the aortic valve, 3D CT images are standard while 4D CT images are preferred for mitral valve treatments. The analysis of the medical images is performed manually by a trained operator using expert software. This analysis takes time and depending on the experience of the operator may entail human error. Moreover, the analysis may be hampered by the image quality and degree of pathology. Artefacts such as motion, noise or signal resonance can omit anatomical details while pathology can obscure exact anatomical measurements. For example, the identification of the mitral valve annulus is difficult for patients with a high degree of mitral annular calcification. The image quality and degree of pathology are not only essential factors for pre-operative planning or as a pre-processing step for simulations but also crucial when developing automatic image analysis methods.

When I started in my PhD in 2015, Dr. Philippe Bertrand generously provided a dataset of 10 patients. Each sample in this dataset consisted of a sequence of ultrasound (US) images and associated measurements of the mitral valve. Armed with two methods [20, 21], I started a quest to extract anatomical features - automatically, because, it was foreseen that my research could automatise the image analysis component within FEops. After a few successes, we had a meeting with some colleagues from the University of Leuven, who explained to us their image analysis tool. This tool helped us to analyse each frame of the image sequence separately and identify anatomical features. When moving to the next frame in the image sequence, all previously identified anatomical features were not taken into account. We quickly realised that if we were going to automatise the image analysis component, we needed to contain, use and include as much data as possible. It was at that moment that the research domain of this PhD shifted towards machine learning and deep learning. One of the first realisations of entering the domain of machine learning is that we had to abandon this dataset because it was too small, plus the interest of FEops moved towards another imaging modality.

Machine learning approaches learn to perform a task by using features that describe the data. Unfortunately, given the high dimensionality of the medical images used in the field of aortic and mitral transcatheter-solutions, two problems may occur. The first problem is the curse of dimensionality. When the number of dimensions of the data is high, many of the machine learning algorithms may not find a proper solution or even fail [22]. For example, predicting a point in a 3D CT image becomes a needle in a haystack problem: the search space is enormous compared to the solution. The second problem concerns feature engineering. Classification methods, such as logistic regression, use features that describe the data. For example, is the patient a smoker or not. The quality of these features is determined by the ability to support the classification task. If the feature is not relevant to the classification, it will not benefit the classification and may even contribute to the curse of dimensionality. However, how to determine the quality of a feature? Also, how do we know when we have

enough features?

## 1.5 Introducing deep learning

Deep learning deals with the two problems mentioned above. Even when analysing high-dimensional data, deep learning will be able to model the data adequately because it can draw non-linear, complex-shaped decision boundaries. Deep learning follows the general principles of machine learning which is that it can learn from large quantities of data and generalise to previously unseen data. However, in order to develop a high performing machine learning model, a rich collection of features must be available. As previously stated, engineering these features by hand may entail difficulty. Deep learning skips this step and learns these features by itself.

Deep learning has revolutionised image analysis in the past decade. It re-emerged in 2012, when a convolutional neural network (CNN) won a global image recognition competition by a large margin [23]. It did not take long before deep learning was introduced in the biomedical domain. In 2015, a two-dimensional (2D) CNN model called U-Net outperformed other biomedical segmentation algorithms in two different competitions [24]. A 3D CNN model called V-Net was proposed one year later and was applied to segment the prostate from magnetic resonance (MR) images [25]. The evolution did not stop here, and in 2018, Gibson *et al.* presented a 3D CNN model called DenseVNet. This model was capable of segmenting several organs from 3D CT images [26].

In this thesis, we present automatic methods that use deep learning to perform cardiovascular image analysis. Given the expected growth of the number of patients, the presented methods may provide powerful tools to enable efficient and accurate cardiovascular medical image pipelines.

## 1.6 Conclusion

In this introduction, I presented FEops, the company where I conducted this PhD research. After which, TAVR and transcatheter mitral solutions were introduced with a strong emphasis on the pivotal role of medical image analysis. Finally, I touched on deep learning and the potential it provides to the field of (medical) image analysis.

## 1.7 Research contributions

In the following sections, I present an overview of the research contributions in this thesis. These sections can be categorised into classification, segmentation and landmark detection.

### 1.7.1 Aortic valve morphology classification

In this study, we hypothesised that deep learning could be used to classify tricuspid and bicuspid aortic valve (BAV) patients from MDCT images. BAV is the most

common congenital heart disorder in adults. The presence of BAV is a crucial factor when selecting the proper treatment for patients with symptomatic severe aortic stenosis who are at increased risk for surgery. Experts claim that it is difficult to classify this pathology, as the leaflet calcification may prevent proper classification and the imaging modality may not reveal the details of the pathology.

This retrospective study was performed on patients who had been referred for TAVR. MDCT imaging was used to label the aortic valve morphology as tricuspid or bicuspid, and a 3D CNN was developed to classify the MDCT images. The cross-validation of the dataset showed that a trained 3D CNN had an excellent ability to classify aortic valve morphology from MDCT imaging correctly, with a high degree of both sensitivity (100%) and specificity (92%). These results are very favourable, especially given the relatively small dataset on which the model was trained. Given these results, it is foreseeable that a tool such as this could be used by clinicians to identify BAV.

## 1.7.2 Segmentation

### 1.7.2.1 Enabling automated device size selection for TAVR

Different measurements of the aortic annulus are crucial during the decision-making process prior to a TAVR procedure. For example, the area or perimeter of the measured aortic annulus determines (among other measurements) the final prosthesis device size. The aortic annulus is measured on the aortic annular plane (AAP); a two-dimensional (2D) plane defined by the lowest hinge points of the aortic valve's leaflets.

In this retrospective study, the images and expert data of 473 patients were used to train two CNN models to segment the aortic annulus from multi-resolution MDCT images. The medical images were resampled to two resolutions (1.0 and 0.5 mm), and each model was trained with a different resolution, but the validation used the input of both models. The area and perimeter were derived from the predicted segmentation and were used to select prosthesis sizes of the Edwards Sapien 3 and the Medtronic Evolut device. The validation was performed in two steps using a test dataset of 118 patients. First, the differences between the manually obtained aortic annulus measurements and the automatic predictions were determined and found to be similar to the differences between two independent observers. These differences indicate a satisfying accuracy of the proposed approach. Secondly, the predicted area and perimeters were inserted in the sizing charts of the Edwards Sapien 3 and the Medtronic Evolute device. It was found that the automatically selected device sizes accorded well with the manually determined device sizes. This accordance confirms the adequate model accuracy again. The total analysis time from AAP to prosthesis size is below 1 second, which is much faster than the manual processing time of approximately 10 minutes. Therefore, the proposed method can speed-up the current pre-operative TAVR planning as well as reduce analysis errors, for example, in inexperienced centers.

#### 1.7.2.2 Towards safe and efficient pre-operative planning of transcatheter mitral valve replacement (TMVR)

The detection of the mitral valve annulus is a time-consuming and challenging process. The mitral valve annulus has a saddle-shaped anatomy, which changes in size and shape during the cardiac cycle. From the detected mitral valve annulus, additional measurements can be assessed such as the area, 2D and 3D perimeter, trigone-to-trigone-, septal-lateral- and commissure-to-commissure distance. The mitral valve annulus and the derived measurements thereof are crucial during the TMVR procedure and mitral valve prosthesis development.

This retrospective study used the images and expert data of 71 patients to train deep learning models to segment the mitral valve annulus from volumetric MDCT images. Once segmented, the derived measurements mentioned earlier were computed during a post-processing step. The method was cross-validated due to the limited size of the dataset. The accuracy of the method was confirmed by comparing the predicted and manually derived measurements. The total analysis time of the proposed method is less than 1 second per patient, which is much faster the manual processing time of approximately 25 minutes per patient. The efficiency and accuracy of the proposed method provide the confidence to move towards implementation of this technology in clinical practice. A proposal for such a medical image analysis pipeline was presented.

### 1.7.3 Landmark detection

#### 1.7.3.1 Aortic annular plane and coronary ostia detection

During the pre-operative planning of a TAVR procedure, the MDCT images are used to identify certain anatomical risk factors. For example, the distances from the AAP to the left coronary ostium (LCO) and the right coronary ostium (RCO) are identified. These distances help to understand the risk of coronary obstruction. This obstruction is a potentially life-threatening complication during which the blood flow to a coronary artery is reduced significantly. The AAP is defined by the three lowest hinge points of the aortic valve's leaflets: the left coronary cusp (LCC), the non coronary cusp (NCC) and the right coronary cusp (RCC) (Figure 1.4).

In this retrospective study, the images and expert data of 344 patients were used to train models that could detect the LCC, NCC, RCC, LCO and RCO. The output of three trained CNN models were combined, and the predicted points were obtained after a post-processing step. The validation was performed on a test dataset of 100 patients (excluded from the training phase), and it was shown that the five landmarks could be detected efficiently and accurately with the proposed method. An inter-observer variability study was also conducted, and the difference between the manually and automatically identified landmarks was generally smaller when compared to differences observed between two independent observers. This difference indicates that the suggested approach detects these landmarks within acceptable accuracy. Finally, it was also illustrated that the method allows to accurately determine clinically relevant measurements such as coronary height - automatically. The total analysis time from

MDCT image to the predicted landmarks was less than 1 second, which clearly shows the potential of the proposed method to speed up current pre-operative planning.

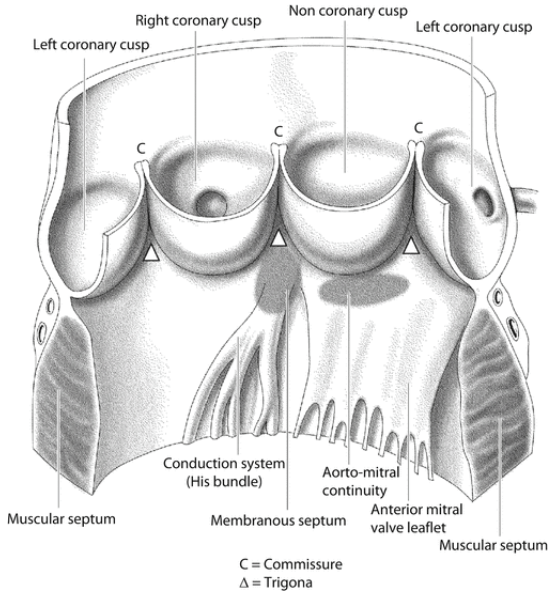



*Figure 1.4: Schematic drawing of the aortic root and components. Image from [27].*

#### 1.7.3.2 Curriculum deep reinforcement learning with different exploration strategies: A feasibility study on cardiac landmark detection

In this retrospective study, a curriculum deep q-learning (DQL) algorithm was used to detect the inferior border of the membranous septum automatically. This border is a known anatomical marker for the atrioventricular conduction path and was defined in our study by three points. The pre-operative assessment of the location of the atrioventricular conduction path is crucial to understand the risk of conduction abnormalities following a TAVR.

The idea behind DQL is that an artificial agent learns to find the points within the medical image. The agent has a limited view and learns to walk towards the target points. During this walk, it tries to maximise some notion of cumulative reward. We applied curriculum learning in order to gradually teach the agent to find the points defining the inferior border of the membranous septum. The method was cross-validated due to the small size of the dataset consisting of 278 patients. The output of this study was three-fold: first, the validation illustrated that the method was accurate and efficient, and an inter-observer variability study confirmed the accuracy. Next, two novel action-selection strategies $\alpha$-decay and action-dropout were introduced and compared to the already established $\epsilon$-decay. Finally, the task of detecting these points is computationally cheap (when compared to similar methods) yet difficult enough to serve as a baseline to experiment with novel DQL techniques.

## 1.8 Structure of this thesis

Since this research is fueled by deep learning, a broader and in-depth introduction of deep learning focussing on image-related tasks is presented in chapter 2. As described above, deep learning has been applied to medical images in order to support pre-operative planning in the field of transcatheter-based structural heart interventions. The clinical background and expected evolution in this domain are presented in chapter 3. Chapters 4, 5 and 6 address methods that perform classification, segmentation and landmark detection in cardiovascular medical images and each contain a state-of-the-art literature overview. The conclusions and future directions of this thesis can be found in chapter 7.

## 1.9 List of realisations

### 1.9.1 Journal publications

1. **Astudillo P.**, Mortier P., Bosmans J., De Backer O., de Jaegere P., Iannaccone F., De Beule M., Dambre J. Automatic detection of the aortic annular plane and coronary ostia from multi-detector computed tomography. *Journal of Interventional Cardiology*, accepted.

2. **Astudillo P.**, Mortier P., Bosmans J., De Backer O., de Jaegere P., De Beule M., Dambre J. Enabling automated device size selection for transcatheter aortic valve implantation. *Journal of Interventional Cardiology*, 2019, https://doi.org/10.1155/2019/3591314.

3. **Astudillo P.**, De Beule M., Dambre J., Mortier P. Towards safe and efficient pre-operative planning of transcatheter mitral valve interventions. *Morphologie*, 2019, https://doi.org/10.1016/j.morpho.2019.09.002.

4. **Astudillo P.**, **Dowling C.**, Brecker S., De Beule M., Mortier P., Dambre J. Classification of aortic valve morphology using convolutional neural networks. *The International Journal of Cardiovascular Imaging*, submitted.

5. Almeida D. F., **Astudillo P.**, Vandermeulen D. 3D Image Volumes From 2D Digitally Reconstructed X-Rays: A Deep Learning Approach In Lower Limb CT-Scans. *Medical Image Analysis*, submitted, 2020.

6. Joana Maria Ribeiro, Thijmen Hokken, **Patricio Astudillo**, Rutger Jan Nuis, MD, Ole de Backer, Giorgia Rocatello, Joost Daemen, Nicolas van Mieghem, Paul Cummins, Matthieu de Beule, Joost Lumens, Nico Bruining, Peter de Jaegere. Artificial Intelligence and Advanced Computer Modelling in Transcatheter Interventions for Structural Heart Disease - Implications for Clinical Practice. *Digital Health Virtual Journal*, 2019

7. Joana Maria Ribeiro, **Patricio Astudillo**, Ole de Backer, Ronak, Rojani, Nicolas M. Van Mieghem, Joost Lumens, Peter Mortier, Paul Cummins, Nico Bruining, E. Boersma, Peter PT de Jaegere. Artificial Intelligence and Tran-

scatheter Interventions for Structural Heart Disease: A glance at the (near) future. *Digital Health Virtual Journal*, 2020.

8. **Astudillo P.**, A. Bhuva. Fully automated deep learning 3D left ventricular echocardiography analysis is generalizable and leads to improved confidence in measuring ejection fraction. *to be submitted*, 2020.

### 1.9.2 Conference publications

1. **Astudillo P.**, Mortier P., De Beule M., wyffels F. Curriculum deep reinforcement learning with different exploration strategies: a feasibility study on cardiac landmark detection *In Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING*, 24-26 February 2020; Valetta, Malta. https://doi.org/10.5220/0008948900370045

2. **Astudillo P.**, Iannaccone F., de Jaegere P., Bosmans J., De Backer O., De Beule M., Mortier P., Dambre J. Detection of the aortic annular plane and coronary ostia from MDCT using deep learning techniques. *ATTICC 2018 - Advanced Techniques and Therapies in Cardiovascular Care*, 11-12 September 2018; Venice, Italy.

### 1.9.3 Abstracts, posters and presentations

1. Almeida D. F., **Astudillo P.**, Vandermeulen D. How accurately can three-dimensional anatomical structures be predicted from a single planar X-ray image? *ESB 2020: $26^{th}$ Congress of the European Society of Biomechanics*; 12-15 July 2020; Milan, Italy; abstract and presentation.

2. **Astudillo P.**, A. Bhuva. Echocardiographic analysis with deep learning: from cardiac cycle classification to cardiac output regression. *$18^{th}$ National Day on Biomedical Engineering: Data Sciences and Healthcare*; 29 November 2019; Brussels, Belgium; poster and abstract.

3. **Astudillo P.**, A. Bhuva. Echocardiographic analysis with deep learning: from cardiac cycle classification to cardiac output regression. *IEEE EMBS Benelux Chapter: Artificial Intelligence in Healthcare*; 28 November 2019; Leuven, Belgium; poster.

4. **Astudillo P.**, **Dowling C.**, Brecker S., De Beule M., Mortier P., Dambre J. Classification of aortic valve morphology using a three-dimensional convolutional neural network. *PCR London Valves 2019*; 17-19 November 2019; London, United Kingdom; poster and abstract.

5. **Astudillo P.** Light-weight algorithm for deep learning architecture evolution applied on image-classification. *Annual International Conference on Computer & Software Engineering*; 22-25 July 2019; Athens, Greece; abstract.

6. **Astudillo P.** Towards Realistic and Tailored Medical Image Generation with Artificial Intelligence. *$17^{th}$ National Day on Biomedical Engineering: Data*

*Sciences and Healthcare*; 30 November 2018; Brussels, Belgium; poster and abstract.

7. **Astudillo P.**, Mortier P., Bosmans J., De Backer O., de Jaegere P., De Beule M., Dambre J. Towards automated TAVI device size selection using artificial intelligence. *5$^{th}$ European Congress on eCardiology and eHealth*; 29-30 October 2018; Moskow, Russia; abstract and presentation.
8. **Astudillo P.** Aortic landmark detection from MDCT images using deep learning. *MUSICARE summer School: "From research to market: new opportunities and challenges in cardiac care"*; 4-8 September 2017; Ghent, Belgium; poster.
9. **Astudillo P.** Efficient & accurate 4D reconstruction of mitral apparatus from echo with deep learning. *MUSICARE summer School: "Multisectoral, multidisciplinary and multiscale features of research in cardiac care"*; 29 September - 2 October 2016; Eindhoven, The Netherlands; poster.
10. **Astudillo P.** Efficient & accurate 4D reconstruction of mitral apparatus from echo. *MUSICARE winter School: "New trends for minimally invasive therapies in cardiac diseases"*; 1-5 March 2016; Lecco, Italy; poster.

### 1.9.4 Awards

1. **Young investigator award** in recognition of the effort and performance of the best presentation. *5$^{th}$ European Congress on eCardiology and eHealth*; 29-30 October, 2018; Moskow, Russia.

### 1.9.5 Patents

1. Method and system for patient-specific virtual percutaneous structural heart intervention. ref: EP18163655

# 2
# Deep learning

*The art of simplicity*
*is the puzzle of complexity*

Douglas Horton

In this chapter, the necessary machine learning and deep learning concepts are introduced to support subsequent chapters. In the first part of this chapter, the general elements of machine learning are explained. These elements are essential to understand concepts such as learning rate, overfitting and others and will return in the rest of this thesis. Since this research was fuelled by deep learning, there is a thorough explanation of deep learning concepts and mechanisms in the final section of this chapter. I refer to the following works for an in-depth overview of machine learning [28, 29] and deep learning [30].

## 2.1 Machine learning

Computer scientist and machine learning pioneer Tom Mitchell defines machine learning as follows; *'A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**'* [28].

In Figure 2.1, an example is presented that shows the interaction between the elements introduced by Mitchell. The task **T** is to classify images of the heart and liver. A learning algorithm trains a model by using the labelled images, which is the experience **E**. During training, a performance measure **P** identifies the accuracy of the model. For example, the model may think that an image of the heart is 80% heart

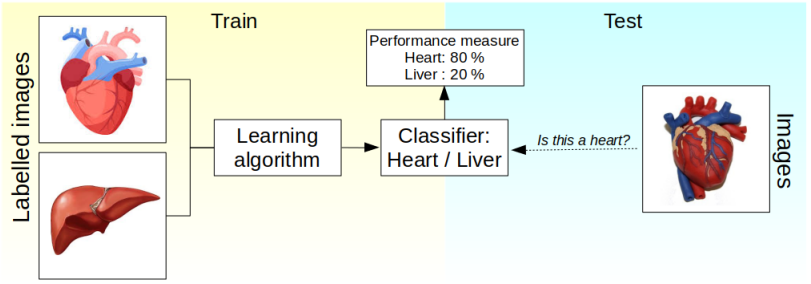



*Figure 2.1: A schematic overview of the interaction between the elements of machine learning.*

and 20% liver. After training, the model can be used to classify previously unseen images during the test phase.

The definition mentioned above is a good starting point to understand the general concepts of machine learning. However, it does not explain how a computer program or learning algorithm learns from data, nor does it explain the organisation of the data, which will be the subject of the following sections.

### 2.1.1 Types of machine learning

In the previous example, a dataset of labelled images was assumed and used to train a model. This type of learning is called supervised learning; where input-output pairs are used to accomplish a task, and the relationship between input and output is known. Supervised learning can be categorised into regression and classification. In the case of regression, a model learns a function that maps the input to a continuous output variable - for example, house price prediction based on the size of the house. In the case of classification, a model learns a function that maps the input to a discrete output variable - for example, cancer prediction based on the size of a tumour.

The second type of learning is called unsupervised learning. This self-organising learning method tries to find patterns in the data without using labels. For example, in the liver and heart classification example, clustering pixels in the images can be an unsupervised task. This clustering could be used to measure blood pools in the images. The final type of learning is reinforcement learning. This area of machine learning focuses on how artificial agents take actions in an environment in order to maximise some notion of cumulative reward.

In this thesis, we will mostly use supervised learning; however, all three types of learning will be used. In the next example, linear regression, logistic regression and neural networks are explained in order to establish some terminology and introduce machine learning concepts.

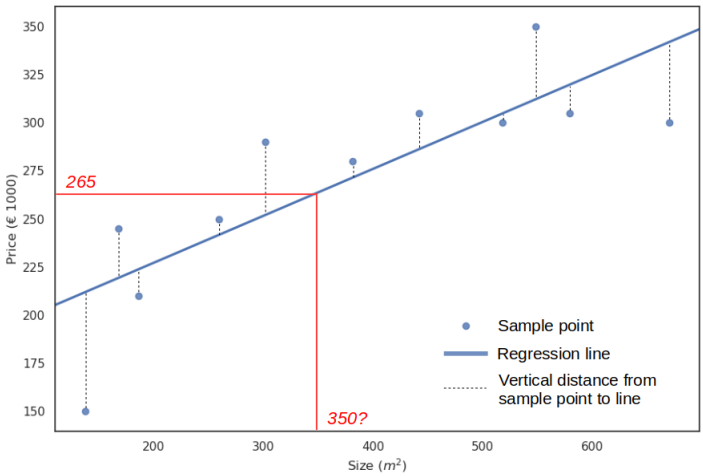



*Figure 2.2: Example of linear regression: house price prediction. The goal is to find the line that best fits the training data. This line can be used to predict house prices of houses which are not in the training set. For example, given a house of 350 $m^2$, the predicted price is 265.000 €.*

### 2.1.2 Linear regression

In this section, linear regression is explained by using a house price prediction example. The size of the houses are the input features of the model (or $x$), and the prices are the output features (or $y$). A pair of size and price is called a training sample (or $(x, y)$), and all training samples form the training dataset. In Figure 2.2, an example is depicted and a generated training set is plotted. As expected, smaller houses cost less than bigger houses. One way to model this trend is to draw a line through the data points. Once we have such a line, we can use this line to predict house prices of houses which are not included in the training dataset.

The goal is to find the line that best fits the training data. This best fit is where the average vertical distance from the plotted training samples to the line will be the least. From the definition of a line, $\hat{y} = w_0 + w_1 x$ where $w_0$ is the y-intercept of the line and $w_1$ is the slope of the line, it is clear that we need to find the values of parameters $w_0$ and $w_1$. In order to evaluate the values of these parameters, the mean squared error (MSE) can be used as a cost function and is defined as

$$MSE(\hat{y}, y) = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$$

where $m$ is the number of predictions, $y$ is the ground truth from the training dataset and $\hat{y}$ is the prediction of the model using the current values of $w_0$ and $w_1$ and the input value $x$ (or the house size in this example).

One method for finding optimal values for the weights is gradient descent. This iterative algorithm tries to find the global minimum of the cost function by showing the entire dataset to the model for several iterations (or epochs). During each iteration,

gradient descent makes small changes to the weights in the direction that reduces the cost. These directions are given by the negative gradient of the cost function. These gradients are the partial derivative of the cost function with respect to the weights. In the case of linear regression, the partial derivatives of the cost function (MSE) with respect to $w_0$ and $w_1$ are computed as follows

$$\frac{\partial MSE}{\partial w_0} = \frac{2}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)$$
$$\frac{\partial MSE}{\partial w_1} = \frac{2}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)x_i$$

and the weights $w_0$ and $w_1$ are updated after each epoch as follows

$$w_0 = w_0 - \alpha\frac{2}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)$$
$$w_1 = w_1 - \alpha\frac{2}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)x_i$$

where $\alpha$ is defined as the learning rate, the most important hyper-parameter of gradient descent.

### 2.1.3 Linear regression with multiple variables

By extending the previous example with an extra input feature, for example, the number of bedrooms, the data can be modelled as follows: $\hat{y} = w_0 + w_1x_1 + w_2x_2$, where $x_1$ and $x_2$ represent the house size and number of bedrooms. Gradient descent can be used to find the optimal values for the parameters $w_0$, $w_1$ and $w_2$ by minimizing MSE with learning rate $\alpha$.

The data range of the house prices will be higher than the number of bedrooms. For example, a house can have a size of 400 $m^2$ while having three bedrooms. This range difference needs to be addressed because gradient descent converges faster on smaller ranges and may overshoot the global minimum (or oscillate) when the input features are highly uneven. Both feature scaling and mean normalisation can be used to address this problem. Feature scaling transforms the range of a feature into a new range of just one and mean normalisation transforms the average of a feature into zero. The advantage of using these two methods is a more stable training process, while a larger learning rate can be used.

### 2.1.4 Logistic regression

In this section, logistic regression will be introduced by using the cancer prediction example. The goal is to classify tumours as malignant or non-malignant by looking at their sizes. In this binary classification problem, the input features are the tumour sizes, and the output can have only two values 0 (non-malignant) or 1 (malignant). Logistic regression uses the same model like linear regression but introduces a sigmoid

function to transform the output of the model into probabilities. The classification is determined by the decision boundary, which is usually set to 0.5. Gradient descent can be used to find the values of the parameters in this model by using the negative log-likelihood (NLL) as cost function which is defined as

$$NLL(\hat{y}, y) = -\sum_{i=1}^{m}(y_i \log(\hat{y_i}) + (1 - y_i) \log(1 - \hat{y_i}))$$

In order to rate a classification model such as logistic regression, the predictions of the model can be compared to the ground truth. A predicted classified sample can be identified as a true positive (TP), a false negative (FN), a false positive (FP) or a true negative (TN) (Table 2.1).

| | **Predicted positives** | **Predicted negatives** |
|---|---|---|
| **Actual positives** | true positive (TP) | false negative (FN) |
| **Actual negatives** | false positive (FP) | true negative (TN) |

*Table 2.1: Confusion matrix to evaluate the outcome of a binary classifier.*

The following ratios are used to identify the performance of a classifier. The positive predictive value (PPV) or *precision* is a fraction of true positive samples over all positive samples. The negative predicted value (NPV) is a fraction of the true negative samples over all negative samples.

$$PPV = \frac{TP}{TP + FP}$$
$$NPV = \frac{TN}{TN + FN}$$

The true positive rate, recall or *sensitivity* is a fraction of true positive samples over the sum of the true positive and false negative samples. The true negative rate or *specificity* is the similar ratio but for the negative samples. Sensitivity and specificity are defined as

$$sensitivity = \frac{TP}{TP + FN}$$
$$specificity = \frac{TN}{FP + TN}$$

The performance of a classifier can also be expressed by computing the f1 score or harmonic average between *precision* and *sensitivity* as follows

$$f1 = \frac{2 * precision * sensitivity}{precision + sensitivity}$$

The f1 score is also known as the Sørensen-Dice coefficient, dice similarity coefficient, or Dice score. In this thesis, the term Dice score will be used. And finally, the accuracy of a classifier can be measured as follows

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

### 2.1.5 Generalisation

The main challenge of machine learning is to create a model that performs properly on new, previously unseen data. This ability is called generalisation, and this is where machine learning differs from optimisation. Training a model can be difficult, and it is essential to understand the generalisation error during training which is why the dataset is split into a training dataset, a validation dataset and a test dataset. This spit enables us to understand the generalisation property of a model during and after training. The training dataset is used to train a model and, the difference between the predicted samples and the ground truth is the training error. The validation dataset is used to evaluate the predictive power of the model during training without altering the weights. The evaluation uses the cost function too, which yields the validation error. The test dataset is used after training the model, and the test error expresses the predictive power of the model to unseen samples.

The training- and test error are related to underfitting and overfitting. Underfitting is the phenomenon where a model is unable to reduce the training error and overfitting describes a significant difference between training- and test error. The best fit is where training- and test error have the lowest value, and the difference between the two errors is small (Figure 2.3).

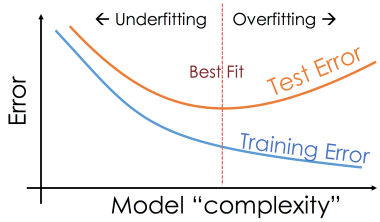



*Figure 2.3: Simplified plot showing the relationship between error vs. model complexity and training- and test error. Image from [30].*

### 2.1.6 Regularisation

Regularisation is defined as *'any modification we make to a learning algorithm that is intended to reduce its generalisation error but not its training error'* [30]. The model's size or capacity is defined by the number of learnable parameters in the model. Whereas the capacity enables the model to represent a wider variety of functions, regularisation can enable a preference towards a specific type of function. Many

regularisation techniques exist, and some of them will be introduced in subsequent sections. One method of regularisation adds a penalty $\Omega$ to the cost function. In the previous example of linear regression, the cost function becomes

$$J = MSE + \lambda\Omega(w)$$

where $w$ are the weights, $\Omega$ is the penalty and $\lambda$ controls the strength of the penalty.

### 2.1.7 Cross-validation

Cross-validation can be used to maximise the amount of data used to train a model and to get an idea of model performance. This approach can be beneficial in case of small datasets. A commonly used method is k-folding, which spits the entire dataset into k subsets. The training and test process is repeated k-times or folds. During each fold, a different subset is used as a test dataset. The remaining subsets are used as the training dataset. K-folding allows using the entire dataset as the test dataset while preserving the separation between training and test dataset (Figure 2.4). If k is equal to the size of the dataset, this method is called leave-one-out.

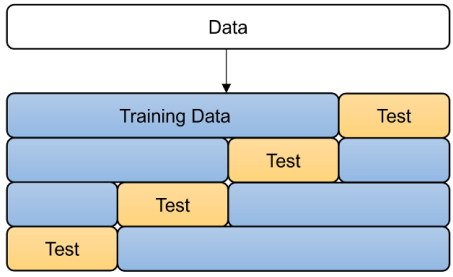



*Figure 2.4: Cross-validation using k-folding (k=4) where the data is split into 4 subsets.*

### 2.1.8 Hyper-parameter tuning

The training of a model uses specific values for the hyper-parameters. The learning rate $\alpha$ is one of the essential hyper-parameters when using gradient descent. Choosing a high learning rate can speed up the training process significantly but may cause gradient descent to miss a local minimum (also known as overshooting). A low learning rate will slow down the convergence speed, but the training process will not miss a local minimum.

There are more hyper-parameters than the learning rate and the regularisation strength controller $\lambda$. In this chapter, many more hyper-parameters will be introduced, which all need appropriate tuning. Hyper-parameter tuning is the process of finding the most optimal values for these parameters. The execution can be performed manually; by testing a different value for a parameter or automatically in which there are two general strategies; grid search and randomised search.

Grid search trains and validates a model for every combination of hyper-parameters. This approach requires a fixed range of values for each parameter. If the most optimal values lie in between this fixed range, the values will not be discovered by executing this grid search once. Randomised search fixes this by choosing random combinations of the hyper-parameters that lie within a specific range.

### 2.1.9 Neural networks

The model behind linear- and logistic regression can be interpreted as a graph or network. In Figure 2.5A, the model of the multiple variable linear regression example is depicted. This network representation has an input layer $L_0$, with inputs $x_0 = 1$ and $x_1 = \textit{house size}$ and $x_2 = \textit{number of bedrooms}$. The arrows represent the multiplication of $x_i$ with the weights $w_i$ and connect $L_0$ to $L_1$, the output layer. In this last layer, the results of the arrows are summed to the output $\hat{y}$. In the example of logistic regression, a sigmoid function is applied to the output $\hat{y}$. The generic term for such a function is the activation function.

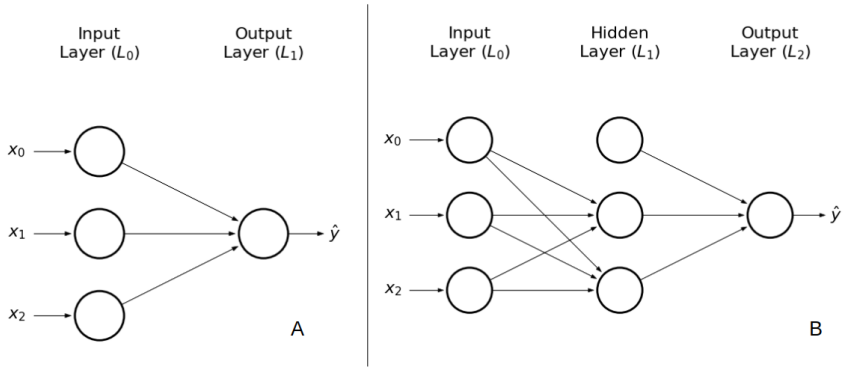



*Figure 2.5: Linear regression as feed forward network (A) and feedforward network with one hidden layer (B).*

It has been shown that the model in Figure 2.5A can derive the binary functions AND and OR, but fails to derive the binary function XOR [30]. This problem can be resolved by transforming the network into a multi-layer feedforward network (or multi-layer perceptron) by adding an extra layer of abstraction in the model. We can visualise this by adding an extra layer in the network as depicted in Figure 2.5B. In this network, there is one hidden layer $L_1$ between the input layer $L_0$ and output layer $L_2$. This extra layer also contains weights (or parameters) and may also contain an activation function. These models are called networks because they compose different functions together in a directed acyclic graph.

Gradient descent optimises the weights of the neural networks in three steps: forward-propagation, back-propagation and weight adjustment. During forward-propagation, input samples flow from the input layer through the hidden layers until the output layer. The cost function computes the difference (or cost) between the

ground truth and the predictions. Back-propagation computes the gradients of the weights with respect to this cost. The negative gradients are used to adjust the weights accordingly with the learning rate $\alpha$.

## 2.2 Deep Learning

The number of hidden layers determines the depth of the network, and when there are five or more hidden layers in a neural network, we enter the domain of deep learning.

In the next sections, essential elements of deep learning are introduced chronologically. This historical ordering was chosen to show that 1), we are standing on the shoulders of giants and 2), deep learning is not some novel technique of the past decade. For the sake of brevity, a short (and incomplete) history of deep learning before 1998 is given as introduction. Afterwards, essential milestones that made deep learning to the field as we know it today are presented. Each milestone is accompanied with the theoretical explanation. Note that there is a strong focus on image related, biomedical deep learning milestones. For an extensive history of deep learning, I refer to the following works [30, 31].

### 2.2.1 A short history of deep learning before 1998

McCulloch and Pitts proposed the first mathematical model for neural networks in 1943 [32]. Seven years later, the father of modern computing, Alan Turing, proposed the imitation game, a method to determine whether a computer can think [33]. This method was renamed and is currently known as the Turing Test. The foundation of deep learning was laid in 1958 when Frank Rosenblatt presented the perceptron model. This model was a system that could recognise *'similarities or identities between patterns of optical, electrical, or tonal information, in a manner which may be closely analogous to the perceptual processes of a biological brain'* [34]. In 1959, the term "machine learning" was first coined by Arthur Samuel, when he presented a checkers program that improved its ability to play by analysing more data [35]. Soon after that, in 1960, Henry J. Kelly published his ideas about control theory, which described the behaviour of systems with input [36]. This work led to the development of back-propagation, which is now used to train neural networks. In 1965, Ivakhnenko and Lapa proposed the first feedforward multi-layer perceptron network [37]. The same authors presented a three-layered model six years later [38] (Figure 2.6A). In 1980, the first convolutional neural network (CNN) was proposed [39] (Figure 2.6B). In 1986, Rumelhart *et al.* discussed back-propagation in great detail [40].

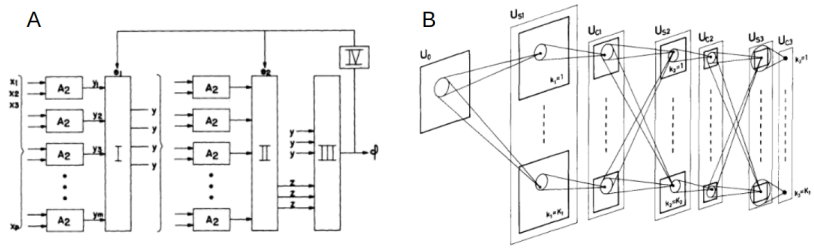



*Figure 2.6: The first deep learning architectures: the neural network proposed in 1971 by Ivakhnenko [38] (A) and the CNN proposed in 1980 by Fukushima [39] (B).*

All the work mentioned above laid the foundations for the deep learning milestones presented in the next sections. These milestones are presented in combination with theoretical aspects related to deep learning.

### 2.2.2 LeNet (1998)

The pioneering work of Yann LeCun and his collaborators classified 32x32 pixel grey-scale images into digits and was applied to recognise hand-written numbers on bank checks [41]. This dataset is currently standardised and known as the MNIST[1] dataset. The seven-layered CNN, LeNet-5 (Figure 2.7), was trained with stochastic gradient descent, a gradient descent algorithm.

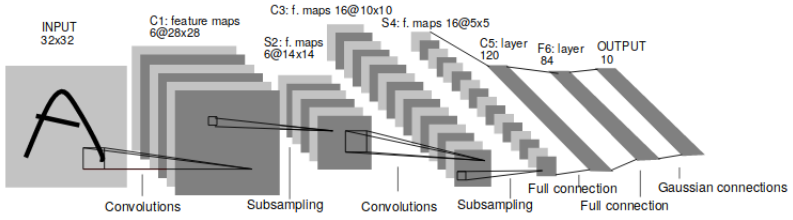



*Figure 2.7: The architecture of LeNet-5 from [41] including convolutional layers, subsampling (or pooling) layers and full connection (or fully connected) layers.*

The architecture of LeNet-5 consisted of convolutional layers, activation functions, pooling layers and fully connected layers. These layers will be explained in the next sections after which the optimisation process, data-augmentation and architectural design will be discussed.

#### 2.2.2.1 Convolutional layers

Convolutional layers contain the functionalities that form the core of CNN models. In this section, the convolutional layer is explained by assuming the analysis of 2D images. The convolutional layer consists of several trainable weights or kernels. Each

[1]Modified National Institute of Standards and Technology database

kernel is smaller than the input but has the same depth. In the example of LeNet-5, the first convolutional layer C1 has six different 5x5x1 kernels (width x height x depth). Each kernel executes the convolution operation during forward-propagation. During this operation, a kernel slides over the input image, and at each position, it computes the dot product between the kernel and the current image patch. Each kernel will produce a 2D feature map that contains the responses of the kernel at every spatial position. The output of the convolutional layer is the concatenation of the produced feature maps from all the kernels. In the example of LeNet-5, the first convolutional layer C1 contains six feature maps as output.

Convolutional layers have the following variables: number of kernels, kernel size, stride, padding and bias. The kernel size defines the shape of the kernel, and thus the receptive field of the layer. The stride defines the step size of the kernel while sliding over the image. Padding defines the border size, and bias determines whether or not to use a learnable vector (the bias) equal to the number of kernels which is added to the output during the forward pass. Figure 2.8 depicts a schematic overview of a convolutional layer and the convolutional operation.

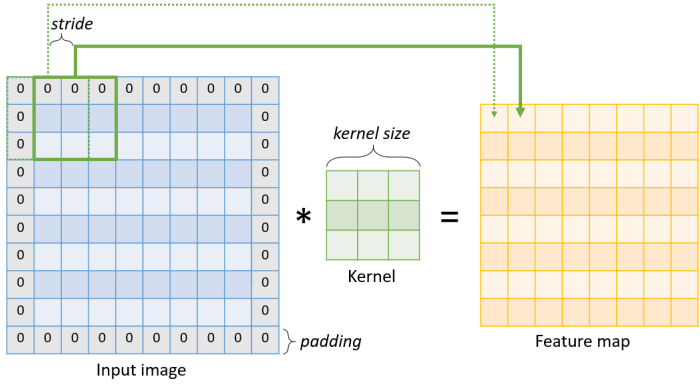



*Figure 2.8: Schematic overview of a convolutional layer with kernel size = 3, stride = 1 and padding = 1.*

Convolutional layers introduce sparse connectivity and parameters sharing. The convolutional layers have sparse connectivity because the kernel size is smaller than the input size. Intuitively, each kernel will recognise some feature that can be detected at any part in the input image. The idea is that if one feature is useful to recognise in some part of the image, it will be useful in other parts of the image. The parameters of the convolutional layer are shared to analyse different parts of the input.

In the example of LeNet-5, the first convolutional layer C1 has six kernels and a stride equal to one. When the convolutional operation starts, the kernel is positioned in the upper left corner. Since there is zero padding and the kernel size is equal to five, the first two pixels in each direction are omitted. This is why the feature maps of C1 have the size of 28x28x6 (Figure 2.7).

#### 2.2.2.2 Activation functions

As previously stated, the sigmoid function is used as the activation function in logistic regression to transform the output of the model into probabilities. Likewise, we can insert an activation function after each layer in a (multi-layer) feedforward network. If the activation function is non-linear, the layer can derive non-linear functions which can assist in obtaining non-linear decision boundaries. Similarly, a (non-linear) activation function can be placed after a convolutional layer.

In the example of LeNet-5, the tanh activation function was used after layers C1, S2, C3, S4, C5 and C6.

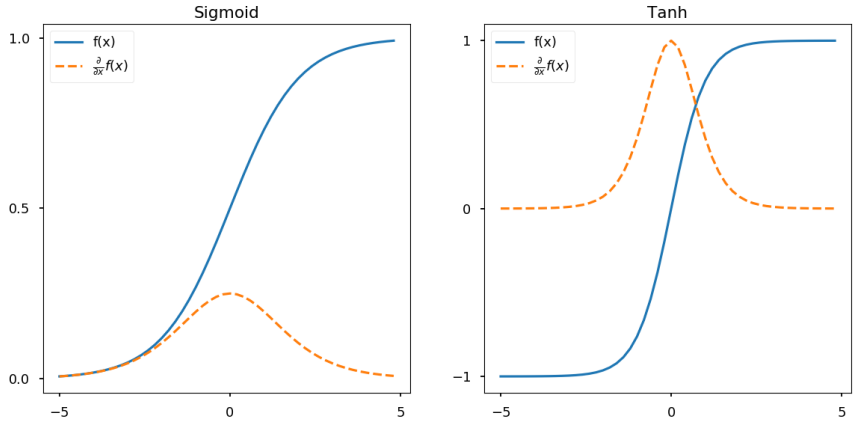



*Figure 2.9: The sigmoid (left) and tanh (right) activation functions and their derivatives.*

During forward-propagation, the output of the activation function is propagated to subsequent layers, whereas during back-propagation, the derivative of this function is used (Figure 2.9). A known problem is the vanishing gradient problem, which means that the gradient becomes smaller and smaller each time it is back-propagated to previous layers. The derivatives of the sigmoid and tanh functions are between 0 and 1, which means that each time the gradient is back-propagated, the signal gets smaller. This problem prevented the construction of deeper networks.

#### 2.2.2.3 Pooling layers

In the example of LeNet-5, average pooling was used in layers S2 and S4. In their architecture, they refer to pooling layers as subsampling layers (Figure 2.7).

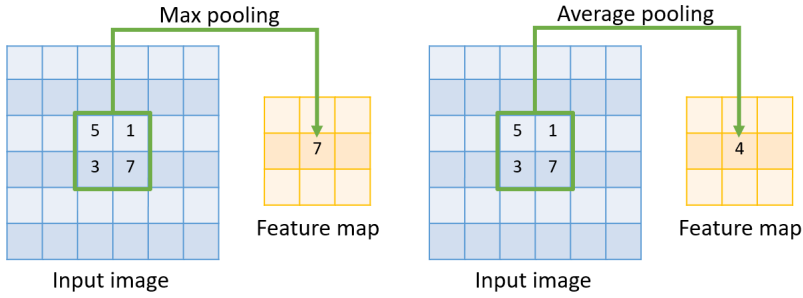



*Figure 2.10: Types of pooling layers: max pooling (left) and average pooling (right)*

The pooling layer contains the operation that reduces the size of the input of that layer. There exist several pooling operations such as max- and average pooling. The pooling layer contains a size parameter which defines the dimensions in which to apply the operation. This size is usually set to 2x2 (in case of 2D image) which reduces the shape on the input by half. In Figure 2.10, max- and average pooling are depicted. It was shown that the pooling operation could also be learned by using convolutional layers with stride equal to the pooling size to reduce the shape of the input [42].

#### 2.2.2.4 Fully connected layer

A fully connected layer or dense layer applies an affine transformation to the input matrix $X$ with shape [batch, features]. This layer contains a weight matrix $W$ with size [features, outputs] and bias vector $\vec{b}$. The affine transformation is defined as $f(X) = W.X + \vec{b}$. The layer connects the input features to the output features as each node of the input feature is connected to each node of the output feature.

In the example of LeNet-5, layer C5 and C6 are fully connected layers. C5 has 400 input features and 120 output features, and layer C6 has 120 input features and 84 output features. The final fully connected layer contains 10 Euclidean radial basis function units which compute the Euclidean distance between the input vector and the parameter vector.

#### 2.2.2.5 Optimisation

As previously mentioned, we can use gradient descent as the iterative optimisation process to gradually alter the weights of a model. If applied to the entire training dataset at once, it computes the correct cost at each epoch. However, this is impossible when the training dataset does not fit in the memory of the computer. In order to overcome this hurdle, gradient descent uses batches of the training dataset. If the batch-size is equal to 1, the optimisation process is called stochastic gradient descent (SGD). This process has the advantage that larger samples can be analysed. However, it entails the disadvantage that the weights of the entire network are modified based on a single experience. This modification may cause the training cost to fluctuate, which increases the total training time. In order to overcome this, SGD with momentum was introduced. This optimiser remembers the previous update and the next update is a

linear combination of the current gradient and the previous update. In the example of LeNet-5, SGD was used to train the model by reducing the MSE cost function.

#### 2.2.2.6 Data augmentation

The training- and validation error are vital to detect under- and overfitting during training. If the model's capacity is large enough to understand the task, the training error will be low. However, when the training dataset is too small, the model may not capture relevant connections between features and target outputs, which will yield a high validation- and test error. This can be resolved by adding more training samples to the training dataset. However, medical images and supporting expert data may be challenging, expensive or painful to acquire. In order to artificially increase the training dataset, a data-augmentation strategy can be used. Such a strategy entails realistic transformations of the samples in the training dataset. In the example of LeNet-5, a robust data-augmentation strategy was presented that added noise and transformations to the training samples, which made the model more robust and prevented overfitting.

#### 2.2.2.7 Architectural design

Deep learning models consist of multiple layers connected in a specific way to each other. The layer types and organisation forms the architecture of the deep learning model. Many types of architectures exist such as the convolutional neural network (CNN), the auto-encoder, the recurrent neural network (RNN), the generative adversarial network (GAN) and many others. The CNN is a model, such as LeNet-5, which uses convolutional layers to encode the image, pooling layers to decrease the size of the image and a stack of fully connected layers to decrease the size of the feature vector into a c-sized vector (where c is the number of classes). A CNN model can be designed following the available literature but can also be evolved (see Appendix B). An auto-encoder is a model that uses an encoder and a decoder to transform an input image into an output image. A variation of such an architecture will be explained in section 2.2.10 of this chapter. RNN is a model which can be used to process sequential data such as text, video or 4D ultrasound images of the heart (see Appendix C). A GAN uses two models: a generating model and a CNN model (see Appendix A).

### 2.2.3 ReLU (2011)

The next deep learning milestone in this chapter is the rectified linear unit (ReLU) activation function or ramp function. This novel non-linear activation function was proposed to overcome the vanishing gradient problem [43]. The derivative of the ReLU function is either 0 or 1. Therefore, the gradient will not vanish due to a decreasing gradient. The function is non-differentiable at 0, but this is usually taken care of by returning 0 or 1 in the used framework. Since ReLU is one-sided, this function may become inactive after a while for any given input. This problem is called the dying ReLU problem and was resolved by introducing leaky rectified linear unit (Leaky

ReLU): a non-linear activation function similar to ReLU which allowed a small positive slope for $x < 0$ (Figure 2.11).

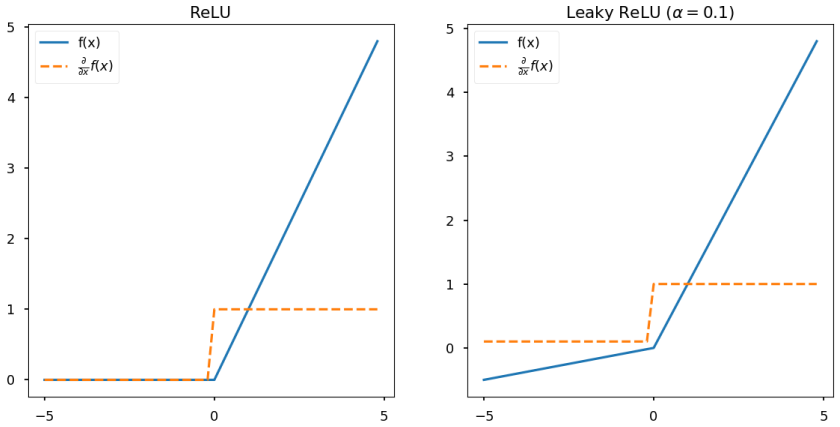



*Figure 2.11: The ReLU (left) and Leaky ReLU (right) activation functions and their derivatives.*

### 2.2.4 Dropout (2012)

Dropout is a regularisation technique for improving neural networks by reducing overfitting. Deep learning models trained with gradient descent build up learned features that work for the training data but may not work for unseen data. Random dropout breaks up these learned features by making the presence of any particular hidden unit unreliable.

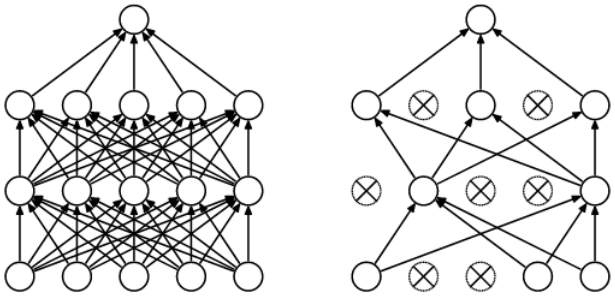



*Figure 2.12: Dropout layer. Image from [44].*

This technique was found to improve the performance of neural nets in a wide variety of application domains including object classification, digit recognition, speech recognition, document classification and analysis is of computational biology data [44, 45].

During the training at each iteration, a percentage ($p$) of randomly selected weights are temporary set to zero in order to reduce the throughput of the network. Only the reduced network is trained and forces the network to look for other solutions. Afterwards, the removed nodes are reinserted into the network with their original weights (Figure 2.12).

### 2.2.5 AlexNet (2012)

The next deep learning milestone is considered the beginning of the era of deep learning in its decennium. Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton created a deep CNN model that won the ImageNet Large-Scale Visual Recognition Challenge of 2012. ImageNet is an extensive and free database containing images and their associated label and was created by Fei-Fei Li in 2009. AlexNet won the competition as it was significantly better than the other competitors [23].

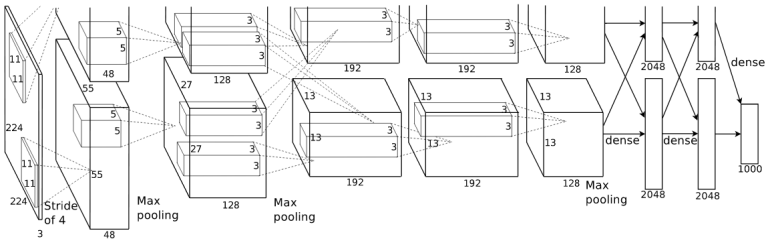



*Figure 2.13: The architecture of AlexNet. Image from [23].*

The AlexNet architecture consists of five convolutional layers, max-pooling layers, dropout layers, and three fully-connected layers and ended by a 1000-softmax activation function. The softmax activation function transforms the output of the layer into probabilities, and the notation of n-softmax refers to the number of classes of the classification task. Figure 2.13 depicts an abstract overview of this architecture. The network was split into two streams because the authors needed to train the model on two GPUs due to memory issues. This architecture was equipped with the ReLU activation function, which enabled faster training when compared to the *'traditional saturating neuron models'* that would use a sigmoid or tanh activation function. The reported total training time for the network was five to six days. The authors also used a data-augmentation strategy that included techniques such as image translations, horizontal reflections, and patch extractions. The authors used SGD with a batch-size of 128 as optimiser to train the model.

### 2.2.6 Random orthogonal matrices (2013)

The initialisation strategy of the weights of fully-connected layers and convolutional layers is another hyper-parameter which needs to be tuned. Unfortunately, there is no superior initialisation scheme. The only certainty we have is that the weights of these layers need to be initialised with random values in order to break the symmetry.

If the weights of a model are initialised with a certain number, for example 42, all weights will be adjusted by gradient descent in the same way, and the model would not learn any different features.

In this milestone, a weight initialisation strategy was proposed that uses random orthogonal matrices (ROM) [46]. This strategy helps to overcome the vanishing and exploding gradient at the beginning of the training due to the norm-preserving property of orthogonal matrices. Moreover, since orthogonal matrices have orthonormal rows, the weights in fully-connected layers and the kernels in convolutional layers will learn distinct features. In the rest of this thesis, ROM will be used to refer to this strategy.

### 2.2.7 VGGNet (2014)

Simonyan and Zisserman explored the depth of the deep learning architectures in [47]. The authors experimented with architectures consisting of 11, 13, 16 and 19 layers. The model employed several convolutional layers with 3x3 kernels followed by a max-pooling layer. The, currently standard, idea of doubling the number of features after each max-pooling layer was first described in their work. In Figure 2.14, the model with 16 layers is depicted.

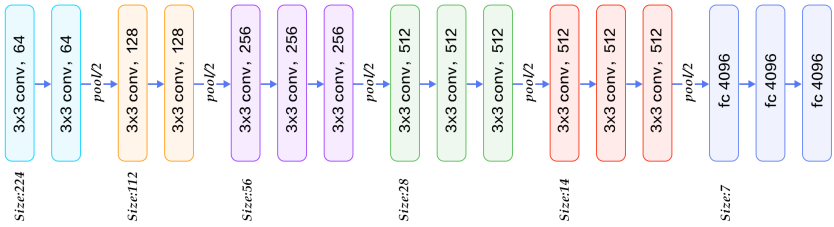



*Figure 2.14: The architecture of VGG-16. Image from [47].*

### 2.2.8 Adam (2014)

Adaptive moment estimation (Adam) is a gradient descent algorithm that computes individual learning rates by using the running averages of both the gradients and the second moments of the gradients [48]. The Adam optimiser is considered faster than SGD because of the combination of momentum and adapted learning rates which may overcome saddle points or local minima. This optimiser was used for most of the research described in this thesis.

### 2.2.9 Batch normalisation (2015)

The idea behind batch normalisation is similar to the idea of normalising the dataset, as discussed in section 2.1.3. In the case of multi-layer feedforward networks or CNN models such as VGG-16, changes in shallow layers will affect deeper layers. In order to prevent this, batch normalisation ensures that the input of each layer is normalised to zero mean and unit variance (by using the mean and standard deviation of the current mini-batch). As a consequence of using a normalised input at each layer,

regardless of its depth in the model, batch normalisation enables the construction of deeper networks [49].

### 2.2.10 U-Net (2015)

U-Net was proposed in 2015 as a biomedical image segmentation algorithm and was applied in several segmentation competitions which it won by a large margin [24]. This auto-encoder architecture contains two paths: an encoder and a decoder path. The encoder path captures the context of the image and contains five blocks, each stacking two convolutional layers and one max-pooling layer. The decoder path reconstructs the segmented image and uses transposed convolutions in combination with the output of the encoded path to enable precise localisation. The ReLU activation function was used in this model. In Figure 2.15, a schematic overview of the architecture is depicted.

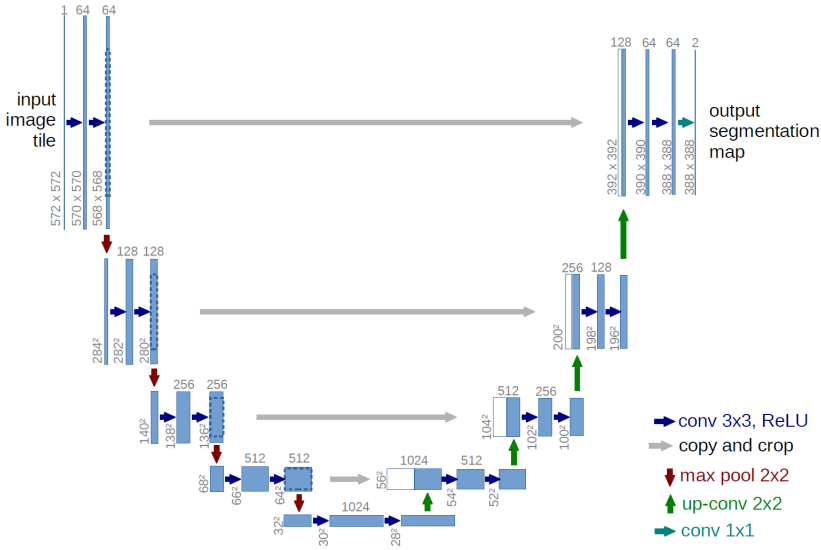



*Figure 2.15: Overview of the U-Net architecture. Image from [24]*

#### 2.2.10.1 Transposed convolutions

Transposed (or fractional) convolutions were used in the U-Net architecture to upscale the input in the decoder path. This layer reuses the convolution layer's jargon in the opposite direction: padding removes pixels from the output, stride results in upsampling, and the filter can be learned or is fixed, for example, bilinear upsampling.

### 2.2.11 Inception (2015)

GoogLeNet [50] was one of the first models which introduced the idea that convolutional layers do not have to be stacked up sequentially. Coming up with the

Inception block, the authors showed that creative structuring of layers could lead to improved performance and computational efficiency (Figure 2.16).

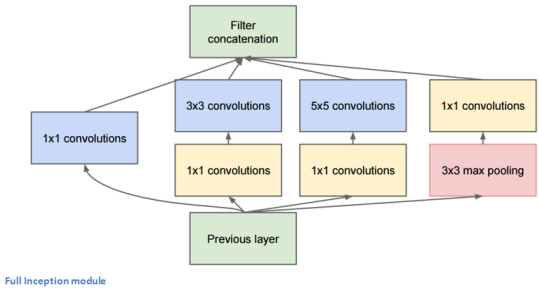



*Figure 2.16: Inception block. Image from [50].*

### 2.2.12 Dilated convolutions (2015)

Dilated convolutions [51] were used to increase the receptive field of the convolutional kernels. Dilated convolutions introduce another parameter to convolutional layers. The dilation rate defines the spacing in the receptive field (Figure 2.17).

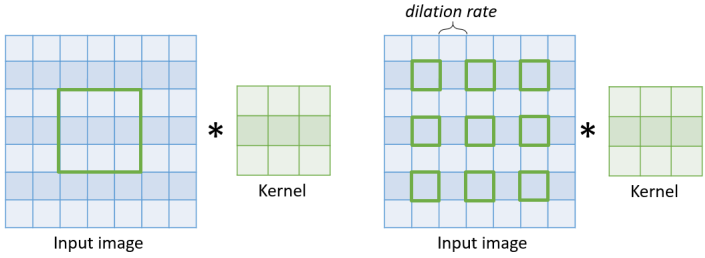



*Figure 2.17: Convolutional layer with kernel size = 3 (left) and dilated convolutional layer with kernel size 3 and dilation rate = 1 (right).*

### 2.2.13 ResNet (2016)

ResNet is a CNN model that proposed the residual block which enabled deeper networks [52]. The idea behind the residual block is the shortcut connection. The input $x$ flows through a conv-relu-conv sequence which results in $f(x)$ and is added to the original input $g(x) = f(x) + x$. The residual block is trying to learn the true output, $g(x)$ but since there is an identity connection, the layers learn the residual. Residual blocks can propagate larger gradients during the back-propagation to the initial layers of the model. This means that the initial layers could learn as fast as the final layers of the model and thus enable deeper networks (Figure 2.18).

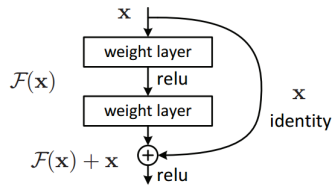



*Figure 2.18: Residual block. Image from [52].*

### 2.2.14 V-Net (2016)

V-Net is a biomedical volumetric segmentation model and was applied to segment MR images of the prostate [25]. The auto-encoder architecture contains two paths similar to U-Net: an encoder- and a decoder path. The encoder path is divided into four stages that operate at different sizes of the input volume. Each stage contains one to three convolutional layers and a pooling layer which reduces the size of the image of the next stage. The decoder path is also divided into four stages, where each stage contains a transposed convolution layer to increase the size of the input, followed by a concatenation of fine-grained features from the encoder stage, followed by one to three convolutional layers. All stages of both encoder and decoder, use a residual block (Figure 2.19).

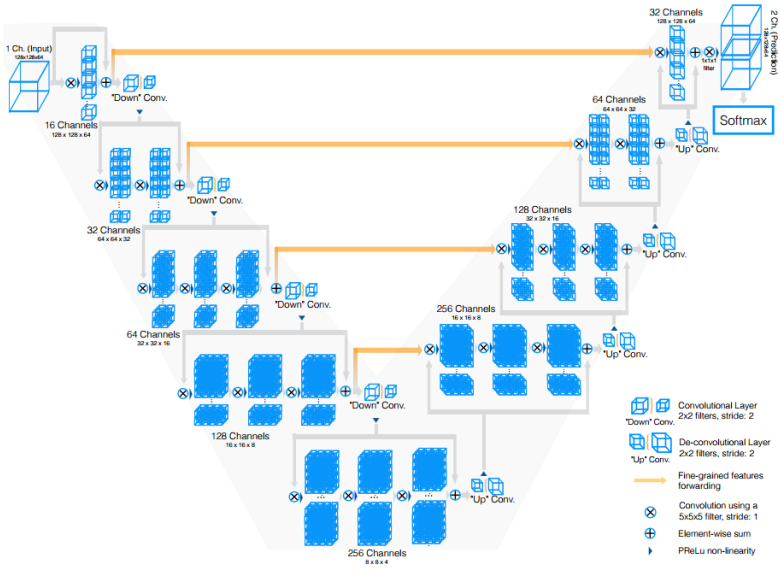



*Figure 2.19: Overview of the V-Net architecture. Image generously provided by Fausto Milletari (first author of [25]).*

#### 2.2.14.1 Dice coefficient cost function

The authors of V-Net used a novel cost function to train the model: the Dice score. Often when segmenting biomedical images, the region of interest (ROI) is small compared to the image, for example, when segmenting the aortic root from a full-body scan. This imbalance is difficult to resolve when training a deep learning model. One possible way of resolving this is by applying more weights to the voxels that define the ROI. However, this may become another hyper-parameter to tune. Another approach is to choose the Dice score as the cost function. Since the Dice score is a metric that measures the amount of overlap between images, the imbalance between the ROI and available search space does not need re-weighting. The Dice score was implemented as follows

$$D(\hat{y}, y) = 1. - \frac{2 * \sum_{i=1}^{N} (\hat{y}_i y_i)}{\sum_{i=1}^{N} (\hat{y}_i^2) + \sum_{i=1}^{N} (y_i^2) + \epsilon}$$

where $N$ is the number of voxels inside the input image and $\epsilon$ is a small value to avoid division by zero. The intersection in the nominator can be written as a sum because the masks are one-hot encoded vectors, a vector of zeros with a single one at the location of the class it represents.

### 2.2.15 DenseNet (2017)

After the introduction of residual connections in ResNet, it was clear that shortcut connections were powerful tools to create deep networks. DenseNet used the effect of shortcut connections and exploited this further by introducing dense blocks. Several convolution layers are included in such a dense block, and every layer is connected to all subsequent layers [53]. The input of each layer consisted of the output of the previous layer and the original input of the dense block (Figure 2.20).

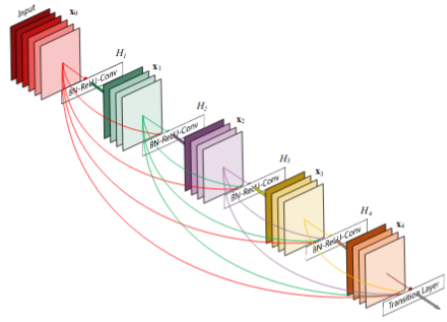



*Figure 2.20: Dense block. Image from [53].*

### 2.2.16 DenseVNet (2018)

DenseVNet was proposed in 2018 as a biomedical volume segmentation algorithm and was applied to identify multiple organs from CT images [26]. The DenseVNet is an auto-encoder and is defined as follows: the encoder path is a sequence of three dense feature stacks whereas the decoder path contains bilinear upsampling to the final output of the network. Dilated convolutions [51] are utilised to increase the receptive field of the convolutional kernels, which may help to detect smaller regions. Dense blocks (or dense feature stacks) are used to encode identity functions which allow effective gradient propagation and reduce the number of activation maps. Finally, batch-wise spatial dropout was introduced, which prevents the computation of activations which would be cancelled by the following dropout layer and thus reduces the memory footprint (Figure 2.21).

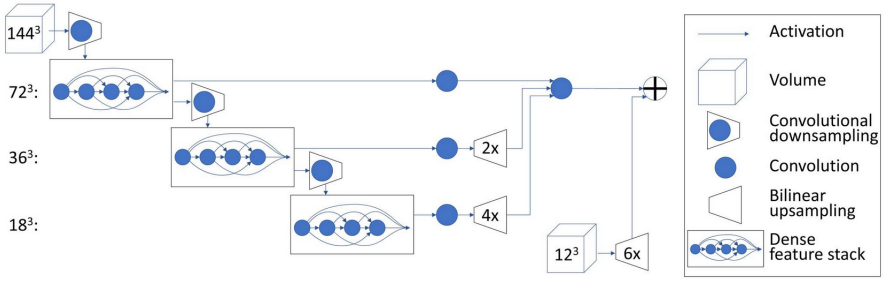



*Figure 2.21: Overview of the DenseVNet architecture. Image from [26].*

## 2.3 Conclusions

This concludes the introductory chapter of deep learning. In the first part of this chapter, the main principles of machine learning were introduced, which are also applicable to deep learning. In the second part, deep learning was presented chronologically with a focus on biomedical image analysis.

# 3
# The human heart

*Sometimes the heart sees*
*what is invisible to the eye.*

H. Jackson Brown Jr.

This chapter begins with the explanation of the anatomy and functioning of the human heart to establish context for the subsequent chapters. Afterwards, a brief overview of valvular diseases and their treatments is presented, followed by an overview of transcatheter-based solutions per valve with a focus on the aortic and mitral valve. This chapter is concluded with a short overview of medical imaging and expert medical image analysis tools.

## 3.1 The human heart

The human heart is a central organ that pumps blood through the blood vessels, which allows nutrients and oxygen to reach the body and removes wastes from the body. This muscular pump is connected to the rest of the cardiovascular system with two circulations: systemic- and pulmonary circulation. The systemic circulation includes the aorta and superior- and inferior vena cava and the pulmonary circulation includes the pulmonary trunk and the left pulmonary veins. The veins are vessels that transport blood to the heart and the arteries are vessels that transport blood to the body. The heart is located between the lungs and in the middle compartment of the chest. The pericardium, a double-layered sac, holds this fist-sized organ in place while allowing it to move as it beats. The wall of the heart consists of three layers: the epicardium, myocardium and endocardium.

The human heart is divided into four chambers: the left atrium (LA) and the right

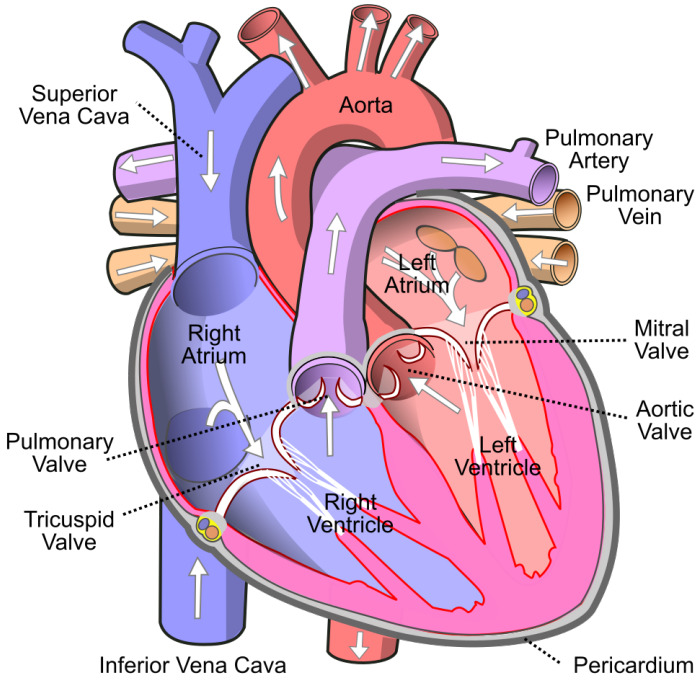



*Figure 3.1: Overview of the blood flow in and around the human heart. Image from Wikipedia.*

atrium (RA) are the top chambers, and the left ventricle (LV) and the right ventricle (RV) are the bottom chambers. Blood flows from the superior- and inferior vena cava, into the RA and is pumped through the tricuspid valve into the RV. The RV pumps blood through the pulmonary valve into the pulmonary trunk towards the lungs. In the lungs, carbon dioxide is released from the blood and oxygen is absorbed. The oxygen-rich blood arrives through the left pulmonary veins into the LA and is pumped through the mitral valve into the LV. In the LV, the blood gets ejected through the aortic valve into the aorta where it gets distributed to the rest of the body (Figure 3.1).

Blood flow is controlled by the cardiac cycle, which is a sequence of systolic and diastolic phases (or pumping and filling phases) of the chambers. At the beginning of the cycle, all the chambers are in diastole, and the atrioventricular valves (the tricuspid- and mitral valve) are open. After this phase, the atrial systole occurs and both atria contract. The blood flows from the atria into the ventricles. Next, during the ventricular systole, the ventricles contract and the blood flows through the opened semilunar valves (pulmonary- and aortic valve) into the great arteries (pulmonary artery and aorta). The atrioventricular valves close to prevent backflow and the atria are relaxed and receive blood. During the next phase, ventricular diastole,

the ventricles relax, and the semilunar valves close to prevent backflow. Finally, all chambers are once again in diastole and the cycle restarts.

The volume of blood within a ventricle at the end of diastole and systole is the end-diastolic volume (EDV) and the end-systolic volume (ESV). Stroke volume (SV) is the difference between EDV and ESV. The ejection fraction (EF) is the SV divided by the EDV and it is the ratio of EDV that is ejected per heart beat.

### 3.1.1 Cardiovascular diseases

Unfortunately, the heart is prone to diseases and cardiovascular diseases are the class of diseases that involve the heart or blood vessels. According to the world health organization (WHO), the number of people dying from cardiovascular diseases will increase from 17.5 million in 2012 to 22.2 million by 2030 [54]. Cardiovascular diseases include coronary artery diseases, stroke, heart failure, rheumatic heart disease, congenital heart disease, valvular heart disease, and others. This thesis concerns transcatheter heart valve interventions, which are treatments for valvular heart diseases.

## 3.2 Valvular heart diseases

Valvular heart diseases occur primarily as a consequence of ageing, but may also be the result of congenital abnormalities or physiologic processes including rheumatic heart disease. The valvular heart diseases can mainly be categorised into stenosis and regurgitation. Stenosis is the stiffening of the tissue of the heart valve typically caused by valvular calcification and causes the narrowing of the valvular orifice which gradually obstructs the outflow of the blood. Regurgitation is the inability of a valve to completely close, which influences the prevention of the backflow of the blood. Stenosis can also result in regurgitation if the thickening of the annulus or leaflets of the valve prevents the valve from closing correctly. The treatment may involve medication but mostly involves surgical valve repair (valvuloplasty) or replacement through the insertion of an artificial heart valve.

### 3.2.1 Treatments

Surgical valve replacement via sternotomy is currently the standard procedure for replacing a diseased valve. This procedure requires general anaesthesia of the patient and an incision is made in the chest to gain access to the heart. The heart is stopped, and a heart-lung (bypass) machine takes over the function of the heart. The diseased heart valve is removed and replaced with a new valve (e.g. a porcine heart valve). The heart is restarted, and the incision in the chest is closed.

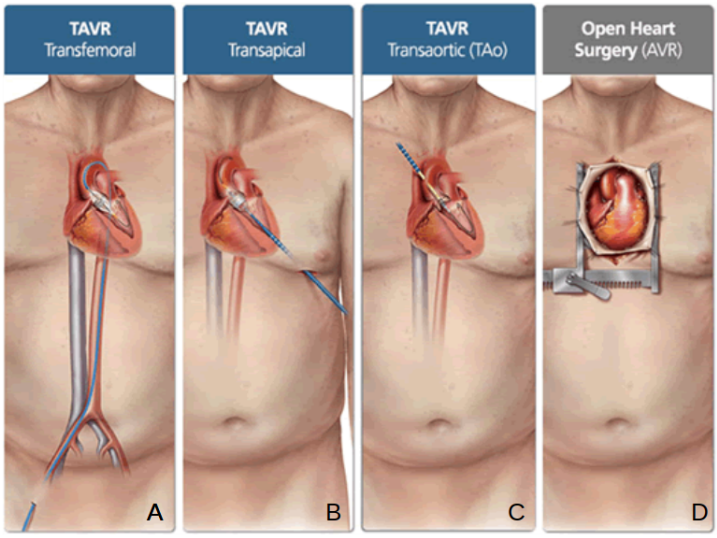



*Figure 3.2: Different ways to access the heart. Transcatheter-based solutions with different entry points: the femoral artery (A), the apex (B) and the aorta (C). And on the right, open heart surgery (D). Image adapted from CVT Surgical Center (www.cvtsc.com/transcatheter-aortic-valve-replacement).*

During open-heart surgery, bone and muscle tissues are pulled and held apart in order to gain access to the heart. Therefore there is a long recovery time and a high post-operative discomfort for the patient. The risk of mortality is high, and therefore, high-risk patients are excluded from this procedure. Mini-sternotomy tries to reduce the recovery time of the patient by using smaller incisions. The risk involved with this procedure is lower when compared to the open-chest procedure, but for high-risk patients, this procedure is still not feasible.

Minimally invasive valve replacement tries to circumvent the long recovery time and post-operative discomfort for the patient by gaining access to the diseased valve through a small incision in the body. A new valve is inserted with a delivery system such as a catheter. After the delivery of the prosthetic valve, the catheter is removed, and the incision is closed. Since the incision and the area of muscle tissue that needs to repair are small, this procedure is considered less invasive than (mini-) sternotomy. Since there is no direct access to the heart, these procedures are also referred to as closed-chest procedures. Different routes to access the heart are depicted in Figure 3.2.

### 3.2.2 The success of transcatheter-based solutions

The transcatheter heart valve interventions were applied to treat the aortic valve, and due to the success in this domain, similar techniques and devices were developed for the other heart valves. As a result, the global market for transcatheter heart valve

treatment devices is predicted to grow from $4 billion in 2018 to $8 billion by 2023[1]. In the next sections, transcatheter-based solutions are discussed per valve.

#### 3.2.2.1 Aortic valve

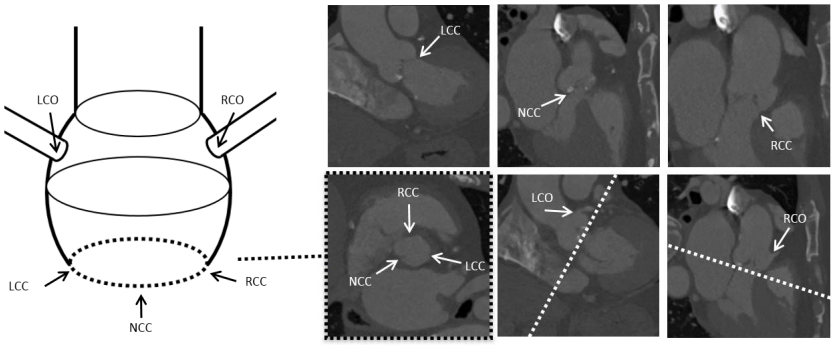



*Figure 3.3: The schematic representation of the anatomy of the aortic root (left). The top three images depict the LCC, the NCC and the RCC (from left to right). The bottom three images show the AAP (dashed line), the LCO and the RCO.*

The aortic valve lies between the left ventricle and the aorta. This semilunar heart valve normally has three cusps or leaflets which prevent the blood from flowing backwards during ventricular diastole. The three cusps are called the left coronary cusp (LCC), the non coronary cusp (NCC) and the right coronary cusp (RCC). The aortic annular plane (AAP) is a plane defined by the three hinge points of these cusps. The orifices of the left- and right coronary artery are located above the left- and right coronary cusps. These orifices are called the left coronary ostium (LCO) and the right coronary ostium (RCO) (Figure 3.3).

Aortic stenosis (AS) is a progressive valvular heart disease that reduces the motion of the aortic leaflets and valve area [3]. The conventional treatment for aortic stenosis is open-chest surgical aortic valve replacement (SAVR) but for patients at high- and intermediate surgical risk, a minimally invasive alternative, called transcatheter aortic valve replacement (TAVR), has become a well-accepted procedure [4–6]. During this minimally invasive alternative procedure, a crimped prosthetic valve is positioned in the aortic root and deployed. The calcified native leaflets are crushed against the aortic wall by expanding the metallic device frame and prosthetic leaflets attached to this frame take over the valve's function. The replacement valve can be delivered via the upper leg (transfemoral) or through the lowest point of the heart (transapical). The delivery of such a prosthesis is depicted in Figure 3.4.

[1] https://www.bccresearch.com/market-research/healthcare/transcatheter-treatment-hlc204a.html

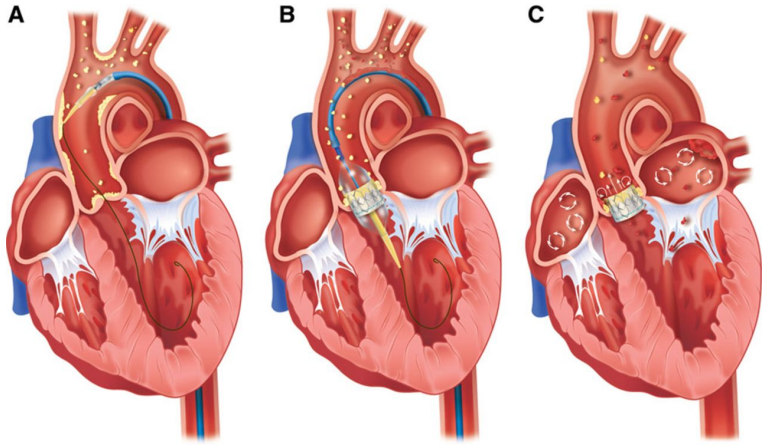



*Figure 3.4: Schematic overview of a TAVR procedure. From left to right: the catheter is positioned through the aorta into the aortic valve (A), a balloon is inflated and deploys the crimped prosthetic valve (B) and the prosthetic leaflets take over the valve's function (C). Image from [55]*

Since TAVR is a closed-chest procedure, the pre-operative planning is image-based. Multidetector computed tomography (MDCT) imaging is the gold standard during the pre-operative planning and allows the heart team to gain insights on the cardiac anatomy and functioning of the aortic valve [7, 56].

The first-in-human procedure started as an investigational technique (or last resort method) for an inoperable patient in 2002 [8]. Due to increasing clinical evidence, TAVR soon became an alternative for SAVR for certain patient groups [4, 57]. The first randomised clinical trial compared treatment with TAVR with the standard procedure for patients with severe aortic stenosis who were not suitable for surgery [9]. The conclusion of this trial[2] was in favour of TAVR. In 2014, a similar non-randomised clinical trial was performed using a Medtronic device with similar conclusions [10]. Finally, a clinical trial showed that TAVR had similar clinical outcomes than the open-heart procedure [11]. These trials established TAVR as an alternative to surgery for patients at high surgical risk. More recently, an investigation was conducted to see whether the patients at intermediate risk could also be candidates for TAVR instead of the surgical approach [12, 13]. Intermediate-risk patients are patients who are estimated to survive 30 days after the procedure [58]. It was concluded that for intermediate-risk patients, TAVR was similar to SAVR concerning the primary end-point of death or disabling stroke. Very recently, clinical data showed that TAVR is at least as good as SAVR in low-risk[3] patients [14, 15].

Currently, there are four clinical trials ongoing to assess the inclusion of low-risk patients for treatment with TAVR over SAVR. All of these trials will add significantly

[2]This trial was made possible by Edwards Lifescience, the company behind the SAPIEN 3, a type of prosthesis for TAVR

[3]Low-risk patients are patients with an STS-PROM score $\leq 3\%$ and absence of co-morbidity that would increase surgical risk.

to the current evidence for TAVR in intermediate- to low-risk patients and will also allow first insights into long-term results on a broad basis [5] (Table 3.1).

| Trial Name | Patients | End year | Code | Publication |
|---|---|---|---|---|
| DEDICATE | 1600 | 2024 | NCT03112980 | No publication |
| Medtronic Evolut TAVR in Low Risk Patients | 1200 | 2026 | NCT02701283 | [14] |
| PARTNER 3 | 1328 | 2027 | NCT02675114 | [15] |
| Nordic NOTION-2 | 992 | 2029 | NCT02825134 | No publication |

*Table 3.1: Overview of the currently ongoing clinical trials to compare TAVR vs SAVR for low risk patients. The code refers to the reference used in the clinicaltrials.gov webpage.*

There will be a significant increase of the number of patients undergoing TAVR in the upcoming years and the reasons for this are clear: younger patients are eligible for TAVR because this minimally invasive procedure includes high-, intermediate- and low-risk patients. The recent change to treat also bicuspid patients with TAVR will also affect the age and number of patients administered [59]. The global TAVR market revenue also reflects these numbers. It was valued at $2.7 billion in 2017 and is expected to reach $8.2 billion by 2025, growing at a compound annual growth rate (CAGR) of 13.8% from 2018 to 2025. The number of cases was 107011 in 2017 and is expected to reach 337778 patients by 2025[4].

#### 3.2.2.2 Mitral valve

The mitral valve resides between the left atrium and the left ventricle. This left atrioventricular valve has two leaflets: the anterior leaflet and the posterior leaflet. The leaflets are connected with the chordae tendineae to the papillary muscles to prevent prolapse into the left atrium during ventricular systole. The mitral valve annulus is a saddle-shaped fibrous ring attached to the leaflets, which changes in shape throughout the cardiac cycle (Figure 3.5).

Mitral valve regurgitation is the most common valvular disease in the western world, affecting 1 in 10 individuals aged 75 and older [16, 17] The prevalence of moderate or severe mitral regurgitation in the United States is approximately 2.5 million and this is expected to reach 5 million by 2030 [60]. Severe mitral valve regurgitation may result in heart failure if untreated. The standard of care for these patients is surgical mitral valve repair or replacement. However, many patients are deemed a too high risk for these highly invasive surgical interventions. Percutaneous mitral valve repair can be a viable alternative for this high-risk patient group [18], but recent data shows that also transcatheter mitral valve replacement (TMVR) is an emerging option [17, 19].

[4]https://www.alliedmarketresearch.com/tavi-market

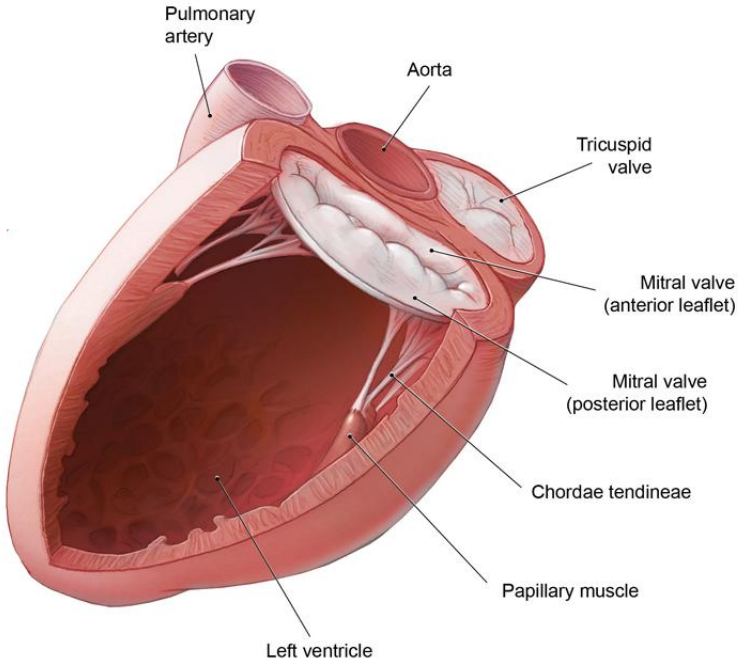



*Figure 3.5: Schematic representation of the details of the mitral valve. Image from Medscape (emedicine.medscape.com).*

The first-in-human native valve TMVR was performed in 2013 in patients with severe mitral annular calcification (MAC) and mitral stenosis [61, 62]. This procedure remains in the infancy of clinical trial testing because of the complexity of the mitral valve apparatus [63, 64]. Currently, the only guideline-recommended transcatheter treatment for mitral valve regurgitation is a valve repair technique using the MitraClip device from Abbott.

First-in-human trans-apical valve-in-valve TMVR was in 2009 in a patient with a stenotic mitral valve prosthesis [65]. After a considerable advancement in the valve-in-valve TMVR field, it is now an established FDA approved therapy [66–69].

#### 3.2.2.3 Tricuspid valve

Tricuspid valve regurgitation is a poorly recognised and frequently overlooked pathology and is usually associated with pulmonary hypertension, a right heart disease caused by inadequate blood flow in the lungs [70]. Transcatheter tricuspid valve repair or replacement is the minimally invasive alternative for a surgical tricuspid valve repair or replacement. The first-in-human transcatheter tricuspid valve repair was performed in 2015 [71] and the first-in-human transcatheter tricuspid valve replacement was also performed in 2015 [72].

#### 3.2.2.4 Pulmonary valve

Transcatheter pulmonary valve replacement is the minimally invasive alternative for a surgical pulmonary valve replacement. The first-in-human procedure was performed in 2000 [73], and this procedure was confirmed in 2015 during a clinical trial [74]. Recently, in 2018, there was a proposal to replace the surgical procedure by the transcatheter approach [75].

The global transcatheter pulmonary valve market was worth $34 million in 2016 and it is projected to reach $51 million in 2023, expanding at a CAGR of 6.2% between 2017 and 2023[5]. The global transcatheter pulmonary valve market is expected to reach a value of $54 million by 2025[6], thanks to the increasing prevalence of various cardiovascular diseases.

#### 3.2.2.5 Transcatheter-based solutions in this thesis

This thesis concerns the medical image analysis prior to a TAVR, TMVR or to facilitate prosthesis development in the mitral valve domain. The image analysis in the tricuspid and pulmonary valve domain was not analysed during this research but are included for completeness reasons and can identify potential future directions.

### 3.2.3 Diagnosis and pre-operative planning

In the previous sections, the different treatments for valvular heart diseases with a strong focus on transcatheter-based solutions were described. The diagnosis, pre-operative planning and treatment of valvular heart disease with minimally invasive techniques are closed-chest and therefore image-based. The image analysis during such procedures is crucial to the success of the procedure. Generally, an entire heart team looks at the images before a procedure to assess the patient's risk, evaluate the vascular access and determine optimal implant size and position. In order to assess the valve accurately, medical images and image analysis software are required. In this section, medical images and expert medical image analysis software are briefly introduced.

#### 3.2.3.1 Medical images

Medical images are required to assess the diseased valve. There are a vast range of image modalities such as magnetic resonance (MR) images, ultrasound (US) images, positron emission tomography (PET) images, and many more. In this research, mainly computed tomography images were analysed.

Computed tomography (CT) is an image recording method that uses combinations of many x-ray measurements to produce an image of the inside of the scanned area without cutting into human tissue[7]. Multidetector computed tomography (MDCT)

[5] http://www.digitaljournal.com/pr/4025571

[6] https://www.grandviewresearch.com/press-release/global-transcatheter-pulmonary-valve-market

[7] The term tomography comes from the Greek words *tomos* (a cut, a slice, or a section) and *graphein* (to write or record).

is the latest type of CT imaging and records the images faster. The x-ray images are recorded along a single axis of rotation and produce virtual slices of the scanned area [76]. These slices are generally stored as a sequence of 2D images and can be stacked into a 3D CT volume. The values of the pixels in a 2D image are expressed in Hounsfield units (HU) which is a linear transformation of the original attenuation coefficient and is defined as

$$HU = 1000 \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}$$

where $\mu_{water}$ and $\mu_{air}$ are the attenuation coefficient of water and air. Hounsfield units enable the selection of specific anatomies with similar attenuation coefficients e.g. soft tissue in contrast CT ranges from +40 to +80 HU or lung tissue ranges from -600 to -400 HU.

CT images are stored in the DICOM format (digital image communication in medicine standard) which also contains meta-information about the patient, recording properties, image information and others. The information about the patient may include name, age and gender. The recording properties may contain information about the recording device manufacturer, the name of the radiologist, the name of the contrast agent used in the study. The image information includes parameters to read the CT image. For example, the number of rows, columns in the 2D slice or pixel spacing which defines the size of a single pixel in the image, or the resolution of the image (Figure 3.6).

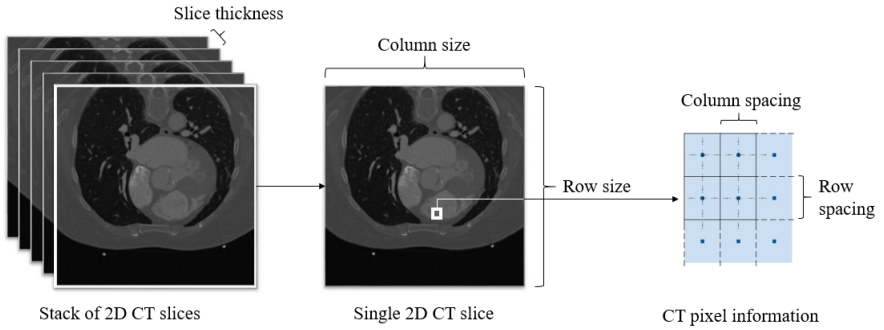



*Figure 3.6: Image and meta-information of CT slices.*

The DICOM format uses a standardised method to include information; however, the information included in the images may differ in each hospital or recording manufacturer. For example, the pixel spacing and slice thickness are parameters that usually differ between two DICOM stacks. Moreover, due to privacy regulations, the images are anonymised when they are shared (for example, for research purposes). This process makes sure that the images of the patient can never be linked to the actual patient and removes some parts of the meta-information of the DICOM images. However, this anonymisation process is also not standardised; some processes only remove the name and study identifier from the DICOM meta-information, and other

processes remove the name, age, gender and other parameters. During this research, different cohorts of patients were analysed. The patients in these cohorts came from multiple centers and all DICOM images were anonymised (see Appendix D for more information about these cohorts).

#### 3.2.3.2 Medical image software

The physician needs to inspect the recorded medical images in order to analyse the anatomical features of the scanned area. Medical image software can read and view these images. There are a few open-source packages such as 3D slicer [8], ITK-SNAP[9] and FSLEyes [10]. These tools allow researchers to view and work with medical images. More professional tools such as Mimics Innovation Suite (Materialise, Leuven, Belgium), 4D Cardio-View (TomTec, Unterschlessheim, Germany) and 3mensio Structural Heart (Pie Medical Imaging, Maastricht, Netherlands) are used in hospitals or companies such as FEops.

The medical image software helps physicians to analyse the medical images, for example, to assess the status of the valvular apparatus and to determine relevant anatomical dimensions. The complexity of the images poses different requirements to the software. A simple image viewer can analyse two-dimensional images. Three-dimensional images, however, should be analysed with three adaptable planes that slice through the volumetric image. Four-dimensional (4D) images (which are 3D images over time), may require a temporal slider to inspect the 3D volumes at each phase of the cardiac cycle.

Manual assessment of the functional behaviour of the valve and the relevant dimensions can be difficult due to the complexity of the valve, disease progression or dimensions of the images. This process typically requires anatomical feature recognition in multiple dimensions and relies on a priori knowledge of the anatomical subject, characteristics of the imaging modality and the tool for visualising the images. It can be a time-consuming job, and the quality depends on the operator's experience. If the quality of the analysis is crucial, a second operator can introduce quality control. This operator receives the data from the previous operator, verifies if the data is correct and alters the data if needed. This second step usually does not take as long as the first step and reduces the operator variability and therefore increases the quality of the analysis.

#### 3.2.3.3 Operators

The users involved in case processing-, case analysts and quality control (QC) analysts go through an extensive training program during which all necessary knowledge is taught. Analysts have to process a set of randomly selected examination cases and pass a predefined set of criteria to obtain a qualification. Process instructions are documented in standard operating procedures. Only after successful qualification,

[8] https://www.slicer.org
[9] http://www.itksnap.org
[10] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes

the analyst obtains access to the processing of the medical images. Training and qualification are governed by an ISO 13485:2016 compliant quality management system.

## 3.3 Conclusion

This chapter covered the introduction of the human heart and medical image analysis in order to provide a foundation for the subsequent chapters. It started with the basic anatomy and functioning of the human heart with a focus on the valves. Next, the valvular diseases and their treatments were introduced, which was followed by an overview of the transcatheter treatment per valve. Since transcatheter-based solutions are closed-chest, the importance of pre-operative medical image analysis was accentuated. Finally, the complexity of this (multi-dimensional) medical image analysis and the potential limitations of medical image software were explained.

# 4
# Classification

*Science is the systematic classification of experience*

George Henry Lewes

Classification is the process of dividing things into groups according to their type. In image analysis, this is the recognition of patterns or objects in images and differentiating them into distinct categories.

This chapter begins with a literature overview of cardiac classification with deep learning. After this overview, we present a deep learning classification method which can differentiate bicuspid- and tricuspid aortic valve patients.

## 4.1 Literature overview of cardiac classification with deep learning

In this section, the available literature of cardiac classification with deep learning is summarised. In order to ensure brevity and relevance reasons, this literature study is limited to the work related to cardiovascular classifications with deep learning. Most of the described work concerns image analysis with a few exceptions where cardiac electrocardiography (ECG) or other signals were analysed. The literature is presented chronologically, and for each year, a summary and detailed overview are given. The summary includes the number of patients, image modality (or signal), model type and model dimensions. The reason for chronologically presenting the literature is to show the evolution of deep learning over the past years. Note that some studies may overlap with the subject of segmentation, which is covered in chapter 5.

### 4.1.1 2015-2016

The published cardiac classification methods were fairly 'simple' in comparison to what was known at that time. For example, architectures contained only convolutional, max pooling and fully-connected layers in combination with the ReLU activation function although dropout and batch normalisation layers were already known. The datasets were relatively small, and efforts were made to balance the positive and negative samples in the training and test phase. Still, there was creative usage of the models, e.g. using the features of a convolutional neural network (CNN) model as features for SVM [77] or using classification to segment a region of interest [78, 79]. Some of the studies already tried to predict clinically relevant measurements from the images [80], which will become a pattern in the following years. Table 4.1 contains a short summary of the discussed studies.

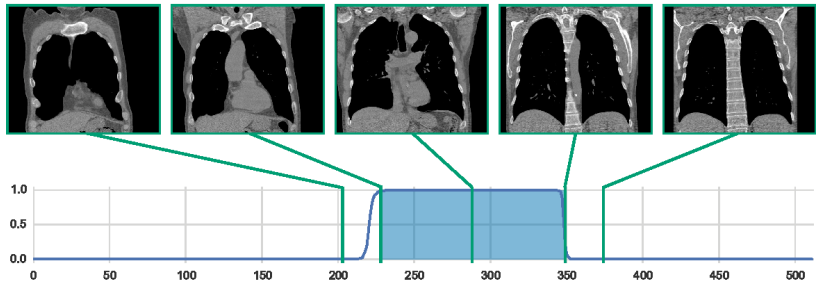
| Authors | Scope | Year | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|---|
| Emad *et al.* [81] | LV localisation | 2015 | 33 | MR | CNN | 2D |
| Vos *et al.* [78] | 3D anatomy localisation | 2016 | 100 | CT | CNN | 2D |
| Zreik *et al.* [79] | LV localisation | 2016 | 60 | CT | CNN | 2D |
| Zhang *et al.* [82] | Apex and basal detection | 2016 | 100 | MR | CNN | 2D |
| Gülsün *et al.* [83] | Coronary centerline pruning | 2016 | 106 | Data | CNN | 1D |
| Moradi *et al.* [77] | Labelling cardiac slices | 2016 | 226 | CT | CNN | 2D |
| Wolterink *et al.* [80] | CAC scoring | 2016 | 250 | CT | CNN | 2D |

*Table 4.1: Summary of cardiac classification with deep learning in 2015-2016.*

#### 4.1.1.1 Detailed literature overview

A new method was proposed by Emad *et al.* to locate the left ventricle (LV) from cardiac magnetic resonance (MR) images by using a CNN model. The architecture of the model consisted of four layers (conv + relu, max pool, conv + relu, max pool) and concluded by a fully-connected layer and a softmax activation function. During training, ten patches (of 20 x 20 pixels) were selected from a short-axis slice, of which one patch contained the LV, and the other nine did not. The dataset consisted of 33 patients, of which 19 were used for training. During testing, a pyramid of scales was used to compensate for the large variety of heart sizes in the dataset. This means that each image was resized into four different scales after which the image was split into patches of 20 x 20 pixels. In order to select the most probable patch that contained the LV, the patch from the scale with the highest probability and least standard deviation was selected. The authors reported a sensitivity of 83.9%, a specificity of 99.0% and an accuracy of 98.6% [81].

In Vos *et al.*, a 3D region of interest (ROI) detector from 2D cardiac non-ECG-gated computed tomography (CT) images was proposed. One CNN model per orthogonal axis was trained to distinguish whether a 2D CT slice was included in the ROI or not. The predictions of the coronal-, sagittal- and axial model yielded a 3D cube specifying the region of interest (Figure 4.1). The trained CNN models demonstrated an excellent discriminatory power to correctly classify ROI (area under the roc curve (AUC) = 0.99). The region of interest was evaluated and the median Dice score of the heart, aortic arch, and descending aorta were 89%, 70%, and 85% [78].




(b) Coronal

*Figure 4.1: Region of interest classification from 2D CT images. The images inside the ROI are classified as one, the others as zero. Clipped image from [78].*

Going one step further, Zreik *et al.* proposed a method to segment the LV from CT images in two steps. The first step used the method described by Vos *et al.* and selected a ROI around the LV. The second step used a CNN model to segment the anatomy within this ROI. The architecture of this model is shown in Figure 4.2. This model was trained patch-wise with a balanced dataset, where the positive samples contained the LV, and the negative samples did not. The accuracy of the automatic segmentation resulted in the average Dice score of 85% and a mean absolute surface distance of 1.1 mm [79].

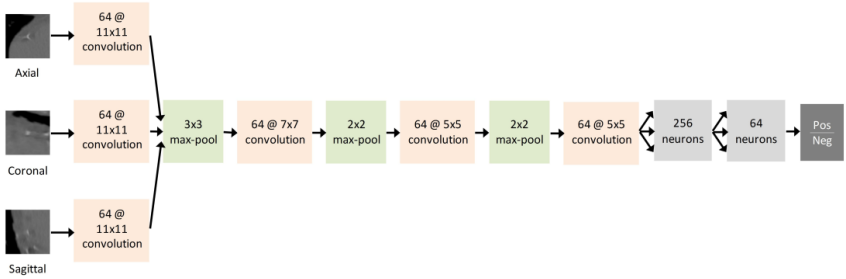



*Figure 4.2: Classification of patches to segment the LV from CT images. Image from [79].*

Zhang *et al.* proposed a new method to detect the basal and apical slices from MR images with two CNN models. The two models used the same architecture with five layers (conv + relu, maxpool, conv + relu, maxpool, conv + relu) and concluded with a fully-connected layer and a softmax activation function. The models were trained to label types of slices such as the apical slice or basal slice from the short-axis slices. Instead of using a patch-wise approach with a fusion strategy (as described by Emad *et al.*), the authors proposed a strategy that used the entire image. They reported the precision and sensitivity of the missing apical slice (81.6% and 88.7%) and the missing basal slice (74.1% and 88.7%) [82].

The main scope of the method by Gülsün *et al.* concerned coronary centerline detection from CT images. The deep learning part of the method was applied after the coronary detection and distinguished true centerlines from false centerlines (or leakages). The one-dimensional (1D) model architecture consisted of four layers (conv + relu, max pool, conv + relu, max pool) and a fully-connected layer followed by a softmax activation function. The 1D data that was used for training came from the centerline detection method and consisted of vessel scale, image intensity, centerline curvature, and other parameters. Their accuracy reporting did not include the classification accuracy [83].

A novel hybrid-learning method was proposed by Moradi *et al.* where a CNN model was applied to label cardiac CT slices. The features from a pre-trained CNN model with five convolutional layers, two fully-connected layers and a softmax activation function were used. The features of the convolutional layers were used in combination with the training dataset in a support vector machine (SVM). The authors concluded that convolutional features from the CNN model yielded higher accuracy than handcrafted features [77].

In Wolterink *et al.*, an automatic coronary artery calcification (CAC) detection method was proposed. The authors used two fully convolutional neural network (FCNN) models with the same architecture and added dropout layers as regulariser. The first model was trained to distinguish the CAC voxels from all negative voxels, and the second model was trained to recognise CAC voxels from CAC-like voxels. The models were trained separately but tested in sequence, where the output of the first model was the input of the second. A patch-wise sampling strategy was applied in order to balance the imbalanced dataset. The authors combined all CNN pairs in an ensemble which assigned 83% of the patients to the same cardiovascular risk category based on the Agatston score[1][80].

### 4.1.2 2017

Similar to the previous period, fairly 'simple' methods were used in comparison to what was known at that time. A notable difference to the previous period is the increase in depth of the models. The datasets that were used remained relatively small. Table 4.2 contains a short summary of the discussed studies.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Acharya *et al.* [85] | Classify heartbeats | 47 | ECG | CNN | 1D |
| Pereira *et al.* [86] | CoA detection in neonates | 151 | US | NN | 2D |
| Santini *et al.* [87] | CAC segmentation | 119 | CT | CNN | 2D |

*Table 4.2: Summary of cardiac classification with deep learning in 2017.*

#### 4.1.2.1 Detailed literature overview

A novel method to classify ECG signals was proposed by Acharya *et al.* This classification method used a CNN model to distinguish five different categories of

[1]This is a quantifier of coronary calcium assessed from cardiac CT [84]

heartbeats. The CNN model consisted of nine layers: conv + relu, max pool, conv + relu, max pool, conv + relu, max pool and concluded by three fully-connected layers and a softmax activation function. To balance the training dataset, the authors added a data augmentation strategy to the training procedure. When the model was trained with the original imbalanced dataset, the validation on the test dataset yielded an accuracy of 90%, and when it was trained with data-augmentation, it yielded an accuracy of 94% [85].

In Pereira *et al.*, a method to detect coarctation of the aorta (CoA), a congenital heart defect, from 2D ultrasound (US) images was proposed. A stacked denoising auto-encoder neural network was trained to detect features from predefined image regions. Similar to Moradi *et al.*, the generated features from the neural networks were used in SVMs, which yielded a total error rate of 12.9% on the test dataset [86].

Extending the study of Wolterink *et al.*, Santini *et al.* proposed a method to detect CAC from 2D CT images. The 10-layered CNN model was trained to classify CAC- and non-CAC image patches. The method was validated on 56 patients and yielded a sensitivity of 91% and a specificity of 95% [87].

### 4.1.3 2018

The significance of the published work increased in 2018, which was reflected by two works published in Nature [88, 89]. All available deep learning methods such as batch-normalisation, generative adversarial network (GAN), recurrent neural network (RNN) and DenseNet were used. Deep learning was not only used as a classifier but also novel diagnostic methods were discovered, e.g. extracting cardiovascular risks from retinal images [89]. There was an enormous increase in dataset size in comparison to earlier years. Table 4.3 contains a short summary of the discussed studies.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Dormer *et al.* [90] | Multi-regional class. | 12 | CT | CNN | 2D |
| Madani *et al.* [88] | Class. cardiac tasks | 267 | US | CNN | 2D |
| Silva *et al.* [91] | Ejection fraction class. | 5600 | US | CNN | 3D |
| Poplin *et al.* [89] | Predicting cardiac risks | 297360 | RI | CNN | 2D |
| Commandeur *et al.* [92] | EAT and TAT class. | 250 | CT | CNN | 2D |
| Dormer *et al.* [90] | Class. of LA | 12 | CT | CNN | 3D |
| Wong *et al.* [93] | Cardiac semantic class. | 371 | CT | CNN | 2D |
| Shadmi *et al.* [94] | Agatston score prediction | 1051 | CT | CNN | 2D |
| Xue *et al.* [95] | Cardiac phase class. | 145 | MR | RNN | 2D |
| Miao and Miao [96] | Coronary disease diagnose | 303 | Data | DNN | 1D |
| Lessmann *et al.* [97] | Calcium scoring | 1744 | CT | CNN | 2D |
| Mazo *et al.* [98] | Class. of cardiac tissues | 6000 | Hist. | CNN | 2D |

*Table 4.3: Summary of cardiac classification with deep learning in 2018*

#### 4.1.3.1 Detailed literature overview

In Dormer *et al.*, a method to classify healthy and unhealthy patients from patches of the left atrium (LA) was proposed. Unhealthy patients were diagnosed with cardiovascular diseases before the image recording. The LA was segmented using

an atlas technique, and patches were generated. The study included a comparison between 2D and 3D patches and concluded that the usage of 3D patches yielded the highest accuracy. The 2D and 3D CNN models were trained and validated with only 10 patients with leave-two-out cross-validation [90].

The significant study by Madani *et al.* evaluated two deep learning classification methods by applying them to a view classification task and a left ventricle hypertrophy (LVH) classification task[2]. The study used 2D US images. The authors compared two approaches: a combined segmentation and classification approach and a GAN approach [99]. The first approach used two famous models: U-Net and VGG-16. U-Net was trained to segment the conic shape of the US images. The predicted segmentation maps were used to select the ROI from the original image. These ROIs were used to train the VGG-16 model to perform the classification task (Figure 4.3). The GAN approach used two models: a discriminator and a generator. The discriminator model used structured blocks of three convolutions and was concluded with two fully-connected layers of size 15 and 1. The former was used to represent the number of classes defined by the classification task, and the latter was used to distinguish generated and real images. For more information about GANs, I refer to appendix A. All models used dropout- and batch normalisation layers. The classifiers were trained to classify 15 categories which corresponded to 15 ultrasound views or 15 classes of hypertrophy. The segmentation + classification approach yielded an accuracy of 94.4% for the view-classification and 91.2% for the LVH classification. The GAN approach was trained with 4% of the images and resulted in an accuracy higher than 80.0% for the view-classification task and 92.3% for the LVH classification [88].

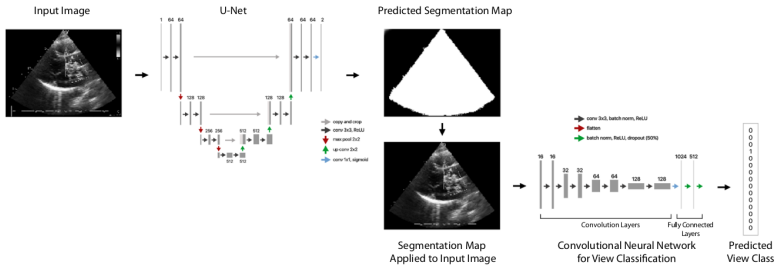



*Figure 4.3: View classification from 2D US images with U-Net and VGG-16. Image from [88].*

Silva *et al.* proposed a novel method to classify ejection fraction (EF) with a CNN model from stacked sequences of 30 2D transthoracic US images. The 3D CNN architecture consisted of residual blocks in combination with dropout layers and L2 regularisation to prevent overfitting. Grid search was applied to detect the optimal values for the hyper-parameters. Ejection fraction was categorised into four classes: $<45\%$, between 45% and 55%, between 55% and 75%, $>75\%$ and an accuracy of 78% was reported [91].

[2]Hypertrophy is an increased volume of organ or tissue due to enlargement of its cells

Another major study was performed by Poplin *et al.*, where cardiovascular risk factors were predicted from retinal images (RI) using the Inception-v3 model. The trained models predicted age, gender, body mass index, smoking status, systolic blood pressure and other cardiovascular risk factors, which was information not thought to be present in RI. The authors trained and evaluated different classifiers: from binary classifiers to classifying all risk factors from RI. The models trained to classify all risk factors yielded the highest test accuracy. This work was ground-breaking, not because of the technical novelties but because the models were able to predict cardiovascular risk factors from retinal images [89].

In Commandeur *et al.*, a novel method to detect the epicardial- and thoracic adipose tissues[3] (EAT and TAT) from CT images was proposed. Two CNN models were used in this study: the first CNN model classified and segmented the CT slices, while the second CNN model segmented the pericardium. The models obtained excellent correlations for both EAT and TAT (0.92 and 0.94). The total analysis time was below 26 seconds per patient [92].

An original method was proposed by Wong *et al.* to classify 2D CT images into semantic cardiac image levels such as aortic root, 4-chamber view and aortic arch. The authors used a 2D segmentation model (inspired by U-Net) where the final concatenated layers were used as input for the 2D classification model (inspired by VGG-16). The scarcity of labelled data was tackled by re-using the weights of a pre-trained segmentation model for the classification task, which achieved an accuracy of 86% [93].

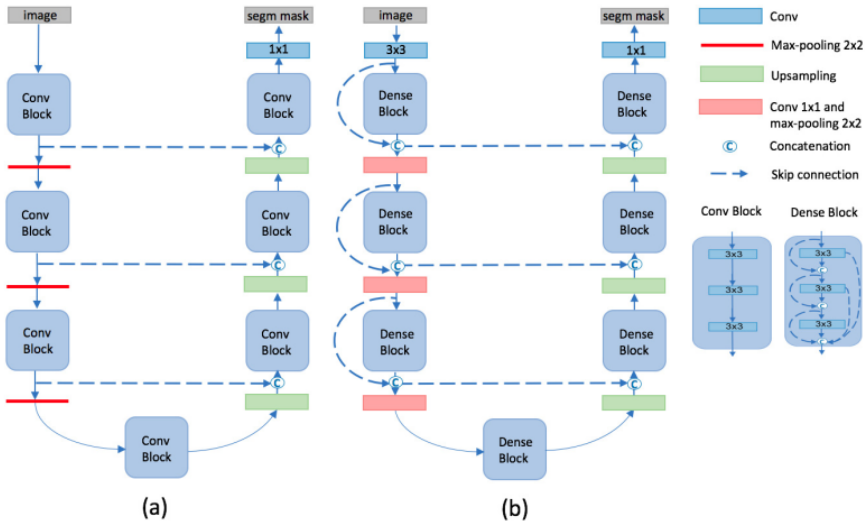



*Figure 4.4: Models used to segment non-contrast chest CT scans to derive the Agatston score. The U-Net model (a) and DenseNet (b). Image from [94].*

[3]Adipose tissues are fat deposits related to coronary artery disease.

Similar to Wolterink *et al.*, a innovative method to predict Agatston score from non-contrast chest CT scans was proposed by Shadmi *et al.* The authors compared U-Net and DenseNet and trained these models to segment the coronary calcification from which the Agatston score was derived during a post-processing step (Figure 4.4). The classification task distinguished five ranges of the Agatston score: 0, 1-10, 11-100, 101-400, >400. DenseNet achieved the highest Pearson correlation coefficient of 0.98 when compared with expert manual annotations [94].

Another method to predict CAC from chest CT scans was proposed by Lessmann *et al.* The method identified patients at increased cardiovascular risk in two steps. The first step, similar to Wolterink *et al.*, a CNN model classified 3D patches into seven classes: left anterior descending artery calcification, left circumflex artery calcification, right/left coronary artery calcification, thoracic aorta calcification, aortic valve calcification, mitral valve calcification and background. In the second step, a model was trained to detect the true positive voxels from the false positive voxels. The coronary artery classification yielded the highest F1 score of 0.89 [97].

Xue *et al.* proposed an original method to quantify the LV from sequences of MR images. The quantification included indices such as area, wall thickness and phase. In order to deal with the temporal aspect of the images, the sequences were encoded with a CNN encoder. These encodings were used in a RNN which was concluded by a softmax activation function. Another RNN model was used to predict LV identifiers such as wall-thickness, area and dimension. The reported error rate for phase classification was 8.2% [95].

A new method to locate five areas of the heart was proposed by Dormer *et al.* The 2D CNN model consisted of six layers (conv + relu, conv + relu, max pool, conv + relu, conv + relu, max pool) and was trained to classify patches of pixels into background, LV, RV, LA, or RA. The method yielded a final AUC of 0.84 [90]. An overview of the workflow is depicted in Figure 4.5.

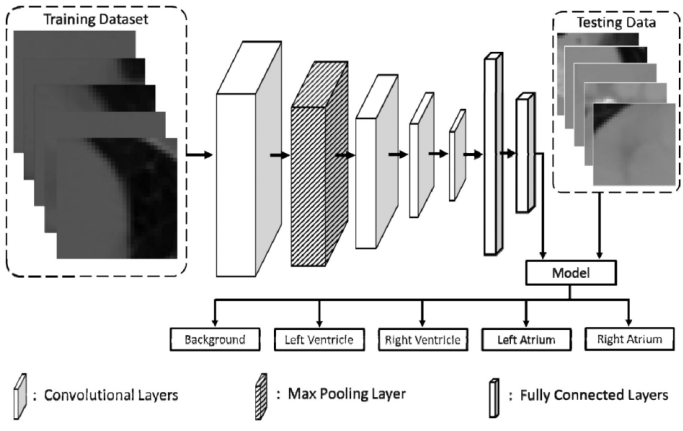



*Figure 4.5: Workflow to detect five regions of the heart from CT images. Image from [90].*

In Miao and Miao, a method to predict coronary heart diseases from patient information such as age, gender, chest pain rating and others was proposed. A deep neural network (DNN) was trained and the validation yielded an accuracy of 83.6%, sensitivity of 93.5%, specificity of 72.8%, precision of 79.1%, F1 score of 0.85 and AUC of 0.89 [96].

Mazo *et al.* performed a comparative study, where CNN models were trained to classify tissue and organs from histology or micro-anatomy images. The authors evaluated four networks: ResNet, VGG-19, VGG-16 and Inception-v1. The models were trained to classify six tissue categories using transfer learning. Overall F1 scores of 0.83, 0.81, 0.82 and 0.75 were obtained for ResNet, VGG-19, VGG-16 and Inception-v1 [98].

### 4.1.4 2019

The peak of 2018 is at the moment of writing not matched. The published work uses established networks (or small alterations thereof) to accomplish accurate results. Still, there is creativity to be found, such as a two-way input network (Gessert *et al.*) or an adapted L1 cost function (Lossau *et al.*). The sizes of the datasets are considerably smaller when compared to 2018. Table 4.4 contains a short summary of the discussed studies.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Lossau *et al.* [101] | Motion artefact recognition | 17 | CT | CNN | 2D |
| Østvik *et al.* [102] | View classification | 500 | US | CNN | 2D |
| Togo *et al.* [103] | CS class. | 85 | PET/CT | CNN | 2D |
| Gessert *et al.* [100] | Plaque detection | 49 | OCT | CNN | 2D |
| Oh *et al.* [104] | Arrhythmia diagnosis | 47 | ECG | CNN | 1D |

*Table 4.4: Summary of cardiac classification with deep learning in 2019*

#### 4.1.4.1 Detailed literature overview

In Lossau *et al.*, a method to recognise and quantify coronary motion artefacts was proposed. A 20-layered ResNet model was trained with coronary CT image-patches to classify motion artefacts. The motion artefact data were obtained from a motion field model, and the motion strength ranged from 0 to 10. The binary classification model considered the values 0 to 1 as 'motion-free' and the values 6 to 10 as 'motion-perturbed'. The same architecture was used as a regression model to predict the real motion strength and trained with an adapted L1 cost function defined as

$$l(y_{regr}, \hat{y}) = \begin{cases} max(0, \hat{y}), & \text{if } y_{regr} = 0 \\ |y_{regr} - \hat{y}|, & \text{if } 0 < y_{regr} < 10 \\ max(0, 10 - \hat{y}), & \text{if } y_{regr} = 10 \end{cases}$$

which improved the training procedure when compared to the L1 cost function. The authors reported accuracy of 93.3% for the classification task and a mean absolute error of 1.12 motion strength for the regression task [101].

A new method was proposed by Østvik *et al.* that classified seven cardiac views in the transthoracic US. The authors compared three CNN classification models: AlexNet, Inception-v3 and a custom network that used inception blocks and residual blocks. The best performing models were saved during training. Their custom model obtained the highest accuracy of 98.3% and the real-time performance of 4.4 ms per frame [102].

Togo *et al.* proposed a novel method to classify cardiac sarcoidosis (CS)[4] and non-CS pixels from positron emission tomography (PET) images. The Inception-v3 network was trained with polar maps generated from the PET images. After training, the high-level features of the model were used in SVM, similar to Moradi *et al.* and Pereira *et al.* This approach yielded a sensitivity and specificity of 94% and 85% [103].

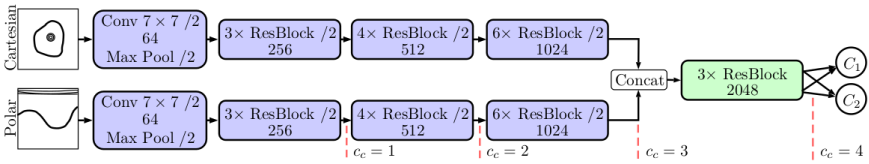



*Figure 4.6: The two-path architecture for simultaneous use of polar and Cartesian representations. Image from [100].*

In Oh *et al.*, a computer-aided diagnostic system to speed up arrhythmia diagnosis was proposed. The 1D U-shaped model was trained with ECG signals to classify five classes: normal sinus beats, atrial premature beats, premature ventricular contractions, left bundle branch block and right bundle branch block. The validation yielded a high overall accuracy of 97.3% [104].

In Gessert *et al.*, a method to detect untreated plaque deposits from intravascular optical coherence tomography (OCT) images was proposed. OCT provides insights into the pathology of coronary arteries by using near-infrared light. The authors compared two CNN models: ResNet50-v2 [52] and DenseNet121 [53], which were trained with both a Cartesian- and a polar dataset. A third model was also evaluated, which had a two-way input path that took both Cartesian and polar images as input (Figure 4.6). The binary classification task tried to distinguish two types of images: plaque vs no plaque. The multi-class classification task tried to classify three classes: calcified plaque, fibrous/lipid plaque and no plaque. In order to overcome the imbalanced dataset, weights were added to the negative log-likelihood cost function, which benefited the class representing calcified plaque. The hyper-parameters were detected with restricted grid search and k-folding (k=3). The authors concluded that the validation of the models trained with Cartesian images yielded the highest F1-score (0.888), which was a surprise considering that the polar images contained a richer representation of the plaque. However, the 2-way path model, which was trained with both image representations yielded a higher F1-score (0.913). The multi-class models

[4]Sarcoidosis is a disease where an abnormal amount of inflammatory cells develop lumps or granulomas.

yielded a lower F1-score (0.833), which could indicate that there was confusion about the types of plaque in the images [100].

### 4.1.5 Meta-analysis of the literature overview

In the above-mentioned literature overview, many medical image classification methods were discussed in great detail; however not compared. In this section, a meta-analysis of the presented classification methods is given. The presented methods ranged from detecting anatomies from a stack of images (such as the apical view) to assisting the risk stratification of a patient. This range can be interpreted in terms of clinical relevance; the detection of the apical slice from MR [82] may assist in the medical image analysis process, whereas the extraction of cardiovascular risk factors from RI [89] may have an impact on how the patient is treated. The presented literature overview can be categorised into four groups of clinical relevance: view detection, anatomical localisation, quantification and risk analysis. These groups will be discussed in the following section.

Classification of medical images can be used to classify an entire image related to the view [77, 82, 88, 93, 102]. These methods can answer questions such as "Is the CT slice the apical slice of the CT stack?" as seen in Zhang *et al.* [82]. This classification may assist the image analysis; however it does not include the location of certain anatomies. The second group encapsulates this localisation of anatomical markers [78, 79, 97, 98]. This group can answer questions such as "Where is the left ventricle in the CT stack?" as seen in Zreik *et al.* [79] and Vos *et al.* [78]. This group contains regions of interest but does not quantify them. This quantification is performed by the third group where a particular clinically relevant quantification is extracted [83, 90, 91, 95, 101]. For example, how much is the ejection fraction in this patient (as seen in Silva *et al.* [91])? This group provides clinical quantification; however, does not assign a particular risk to it. This risk analysis is covered in the last group which analyses the medical images (or signals) and predicts cardiovascular risks [80, 85–87, 89, 92, 94, 96, 100, 103, 104].

The different groups presented above show the potential of deep learning in medical image analysis. However, it should be noted that the studies in these groups introduce novel techniques but do not include some of the latest requirements in AI, mainly related to reproducibility. In order to be able to reproduce the results of a method, the following information should be available. First, an orchestrated reporting of the experimental results which led to the final results should be available [105]. It would be interesting to summarise this information from the presented methods (where available). Next, a detailed overview of the patients in the cohort should be given, or a public image database should be used. Currently, there is no standard method to summarise the medical images from the patient cohorts in order to be able to reproduce the results. Thirdly, a robustness test of the method could verify how robust the method remains during changes of certain parameters specific to the cohort. Examples of such a robustness test may include the age of the patients to see if the method is resilient for younger or older patients. Another example of a robustness

test may include the patient's orientation, image noise or even patient's pathology (sick vs healthy patients). Finally, comparable to regulatory rules, the reproducibility of a method could benefit from an indication for use and limitations of the method. For example, if the model was trained on a cohort of patients between 80 and 90 years old, a limitation could be to exclude younger or older patients. Age is taken as an example here, but this could be extended to race and other demographic parameters. It is essential to indicate that the requirements related to reproducibility have only been introduced recently (and some are still under discussion or development) however they originate from the success of deep learning in cardiovascular classification.

### 4.1.6 Conclusion

This concludes the literature overview of cardiac classification with deep learning. As previously mentioned, deep learning has undergone quite the evolution in the past decade, and this literature overview tried to convey this evolution; from a simple method to locate the LV in 2015 [81] to standard methods for cardiac classification in 2018 [88, 89].

In the next section, a novel method is described that classifies an aortic valve morphology from MDCT images using a 3D CNN model. In the presented literature overview, there are a few studies which use a 3D model for the classification task [90, 91]. Other studies overcome the extra dimension by training a 2D model with a stack of 2D images, yielding a 3D classification [78]. In Figure 4.7, a timeline with the most significant milestones of the presented literature and the classification study in the next section is depicted.

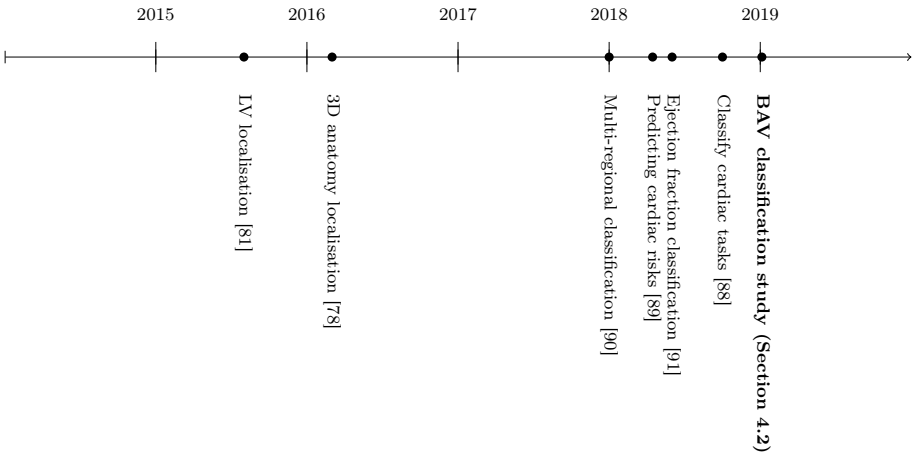



*Figure 4.7: Chronological overview of the most significant milestones of the presented literature and the bicuspid aortic valve (BAV) classification study.*

## 4.2 Aortic valve morphology classification[5]

From the literature overview of this chapter, it was clear that deep learning can be used for many medical image classification tasks. These tasks can range from ROI detection in 3D CT images [78] to the detection of untreated plaque deposits in intravascular OCT images [100]. In the first study, a model classifies an entire image to identify the ROI, whereas, in the second study, the model looked for the presence of small image features to assign a label to the image. In this section, a method is presented that classifies aortic valve morphologies. The method will require both the entire image and the presence of small image features to classify the bicuspid and tricuspid aortic valves from MDCT images using a 3D CNN architecture.

First, some background is provided about bicuspid- and tricuspid aortic valves, followed by a description of the method. Next, the results are presented, and finally, a short discussion is given about the clinical impact of the obtained results.

### 4.2.1 Bicuspid aortic valve

Bicuspid aortic valve (BAV) is the most common congenital cardiac malformation that affects approximately 1% of the general population [106–108]. BAV include a spectrum of deformed aortic valves with only two functional leaflets[6]. The spectrum is classified according to the number of fusions or raphes between the leaflets. Type 0 or the congenital BAV has zero raphes and occurs in about 6% of the BAV patients. Type 1 is characterised by a single raphe and is the most common one (89% of the BAV patients). This type has three subcategories depending on which leaflets the raphe fuses. Type 2 is identified with 2 raphes and occurs in 5% of the BAV patients [109]. In Figure 4.8, a schematic overview of the different BAV types is given.

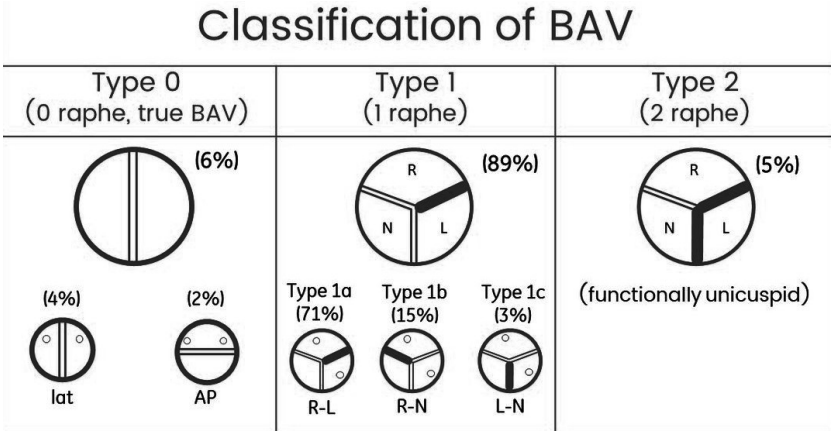



*Figure 4.8: Schematic overview of the different types of bicuspid aortic valves. The classification depends on the number of fusions or raphes between the leaflets. Image from [109].*

[5] in collaboration with Dr. Cameron Dowling

[6] Normally the aortic valve has three leaflets.

People with a BAV are more likely to present with aortic stenosis (AS) at a younger age when compared to people who have a tricuspid aortic valve [110]. Currently the standard treatment is surgical aortic valve replacement (SAVR) [59]. In the early days of transcatheter aortic valve replacement (TAVR), when only high-risk patients were considered candidates for this procedure, the classification of BAV patients during the pre-operative phase was not that important since most of them would have been already operated at an earlier age. However, recent data shows that low-risk patients treated with TAVR have a similar clinical outcome as compared to the surgical approach [14, 15]. This implies that more and younger patients will be considered for TAVR, including the patients with BAV.

Furthermore, for patients undergoing TAVR, the presence of BAV has significant pre- and peri-procedural implications, such as the need to consider the assessment of the supra-annular structure in order to size the transcatheter heart valve prosthesis [111–113], and the consideration of pre-implantation balloon aortic valvuloplasty. The correct identification of BAV on MDCT imaging can be challenging, especially without the aid of specialised 3D reconstruction tools such as the 3mensio Structural Heart version 9.1 (Pie Medical Imaging, Maastricht, Netherlands). As a result, many patients with BAV are not correctly identified. Automated detection of BAV patients would, therefore, be valuable.

In order to automatically detect BAV in MDCT images, the details of the pathology must be explored. On valve level, the one or two leaflets of the aortic valve will be attached, albeit for congenital reasons or due to aortic stenosis. In order to ensure proper pathological classification and not just labelling high amounts of calcium, a dedicated calcium analysis tool was used to quantify the volume of aortic leaflet calcium, with the threshold for calcium set at 700 Hounsfield units (HU). Calcification of the left ventricular outflow tract, sinotubular junction and coronary arteries were excluded from the region of interest. On a broader level, different aortic dilation patterns are described in BAV patients. Aortic dilation occurs due to the malfunctioning of the bicuspid aortic valve. The dilation effect can be measured at the aortic root and arc [114]. In Figure 4.9, examples of aortic dilation are presented.

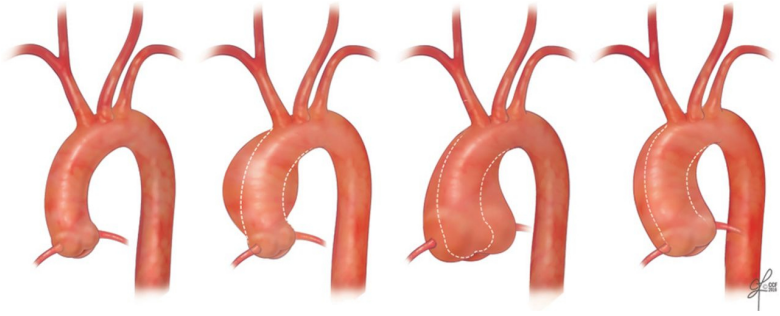


*Figure 4.9: Schematic overview of different aortic dilation patterns in bicuspid aortic valve patients. Image from [114].*

Since this morphology can be detected on the valve/leaflet level and aortic root/arc level, the field of view used to classify medical images becomes an essential factor. On the one hand, low-level details must be included, and on the other hand, a broader view of the aortic apparatus may also play an essential role in the final classification.

## 4.2.2 Materials and methods

### 4.2.2.1 MDCT images

The dataset consisted of ECG-gated contrast-enhanced cardiac MDCT images of 96 patients. The average age of this cohort was $80.8 \pm 7.1$ years, and 50% of the patients were male. An expert manually classified all images. The number of bicuspid (48) and tricuspid (48) patients were kept equal in order to prevent imbalance issues during the training and detecting phase. The patient characteristics of the classification dataset can be found in Table 4.5.

| **Characteristic** | **Value** |
|---|---|
| Number of patients (N) | 96 |
| Age (years) | 80.3±7.4 |
| Male | 48 (50%) |
| STS-PROM score | $4.7 \pm 1.1$ |
| **Aortic valve morphology** | |
| Tricuspid | 48 (50%) |
| Bicuspid | 48 (50%) |

*Table 4.5: Patient characteristics. Values are mean ± standard deviation or n (%). The society of thoracic surgeons predicted risk of mortality (STS-PROM) as a validated predictor of 30-day mortality after cardiac procedures [115].*

### 4.2.2.2 Manual classification

The cardiac MDCT imaging was post-processed in 3mensio Structural Heart version 9.1 (Pie Medical Imaging, Maastricht, Netherlands) using the Aortic Root analysis package. Automatic segmentation of the aortic root was performed, and then the centerline was adjusted manually if deemed necessary. The lowest insertion point of each aortic cusp was identified and used to define the basal annular plane. Automatic short-axis views were then generated. A 'hockey puck' 3D reconstruction of the aortic valve was created, and aortic valve morphology was then classified as tricuspid or bicuspid (Figure 4.10).

| | | |
|---|---|---|
| **Sievers type 0** | | 7 (7.3) |
| | Anterior-posterior | 5 (5.2) |
| | Lateral | 2 (2.1) |
| **Sievers type 1** | | 39 (40.6) |
| | Left-right raphe | 33 (34.4) |
| | Right-non raphe | 6 (6.3) |
| **Sievers type 2** | | 2 (2.1) |
| | Left-right / right-non raphe | 2 (2.1) |

*Table 4.6: Subtype details of the BAV patients in the cohort. Values are number of patients (%).*

The bicuspid valves were further classified using the Sievers classification system [109] and the details are described in Table 4.6. The most frequent BAV classification was a Sievers Type 1 valve with single left-right raphe.

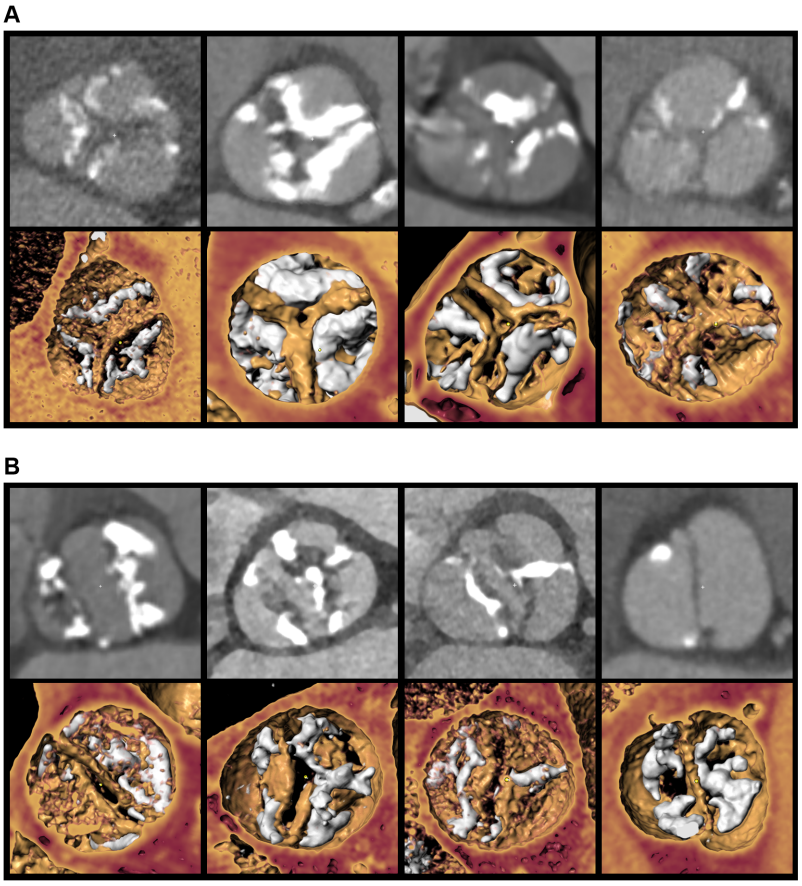



*Figure 4.10: The aortic valve perpendicular plane (gray) and the 3D hockey puck view (coloured) of tricuspid (A) and bicuspid (B) aortic valve patients. Image courtesy of Dr. Cameron Dowling.*

#### 4.2.2.3 Pre-processing

The anatomical differences between tricuspid and bicuspid patients vary from patient to patient but mostly include malformations at the raphes, leaflets, coronaries, valve or aortic arc. Since the MDCT images ranged from full body- to regional scans and entire MDCT image would not fit the memory of the graphical processing unit (GPU), a pre-processing step was required.

The volumetric MDCT images were resampled to five isotropic resolutions (0.25, 0.5, 1.0, 2.0, 3.0 mm) to ensure a homogeneous dataset. These multiple-resolutions were used to increase the amount of information during training and was also observed

in Emad *et al.* [81]. The original Hounsfield units were preserved during the resampling process, which was performed with cubic spline interpolation. In order to focus on the aortic root, all images were oriented around the center of the left-, right-, and non-coronary cusp and clipped to $64^3$ voxel cubes to fit the model's input. The locations of these cusps were identified by an expert in Mimics Innovation Suite 18 (Materialise, Leuven, Belgium). In Figure 4.11, an overview of the different resolutions of the same area in a patient are depicted. It is visible that different resolutions of the same area contain different anatomical information. As a final step, all images were normalised using feature scaling.

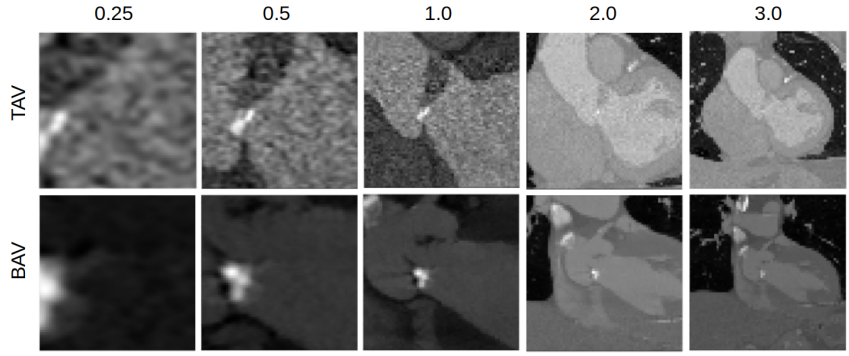



*Figure 4.11: Overview of the exported resolutions for both bicuspid (BAV) and tricuspid (TAV) patients. From left to right: resolutions 0.25, 0.5, 1.0, 2.0 and 3.0 mm. Notice the aortic root dilation of the BAV patient (visible in resolutions 2.0 and 3.0 mm).*

#### 4.2.2.4 Automatic classification

Since there was no 3D CNN deep learning architecture available in the literature at that time, a model was developed to classify tricuspid- and bicuspid patients from a single MDCT image. The model's architecture followed a typical classification architecture which contained two major elements: an encoder and a funnel. The encoder was used to learn anatomical features while step-wise reducing the dimensions of the input. After the encoding step, the entire output was flattened and presented as a feature vector to the second element. The funnel consisted of stacked fully-connected layers which gradually reduced the size of the feature vector and was concluded by an average pooling layer which reduced the size of the feature vector to 1. Finally, a sigmoid activation function transformed the output of the model to probabilities. All tricuspid patients belonged to class 0, and the bicuspid patients to class 1. The decision boundary was set to 0.5 to determine the class of predicted value.

Different experiments were conducted to assess the effect of the amount of anatomical information on the performance of the classifier. The number of resolutions controlled the amount of anatomical information, e.g. a model could be trained with single resolution (1 mm) images or with multiple resolutions (0.25, 0.5, 1.0, 2.0, 3.0 mm). The resolutions were inserted in the channel dimension of the CNN model. In Figure 4.12 an overview of the 3D classification architecture is given.

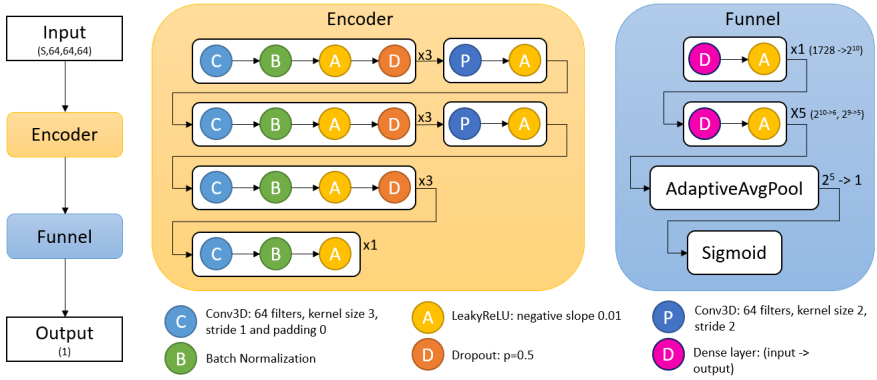



*Figure 4.12: The 3D architecture to classify bicuspid- and tricuspid patients. The input signal flows from the encoder, through the funnel to the output. The number of resolutions **S** is annotated in the Input(**S**, 64, 64, 64).*

**Training details** The Adam optimiser [48] was used to minimise the negative log-likelihood (NLL) cost function with an initial learning rate of 1e-4. A cosine annealing schedule was used to reduce the learning rate towards 1e-6 over time. The weights of the convolutional layers were initialised with random orthogonal matrices with $gain = \sqrt{\frac{2}{1+\alpha^2}}$ where $\alpha$ is equal to the negative slope of the Leaky ReLU activation function ($= 0.01$). Since the cohort size was small, the method was cross-validated using k-folding (with k=4).

### 4.2.3 Results

The trained CNN models demonstrated excellent discriminatory power to correctly classify aortic valve morphology (Figure 4.13). The single resolution (1.0 mm) (AUC=0.95, 95% CI 0.90-1.00, P<0.001), dual resolution (0.5, 1.0 mm) (AUC=0.96, 95% CI 0.92-1.00, P<0.001) and quintuple resolution (0.25, 0.5, 1.0, 2.0, 3.0 mm) (AUC=0.95, 95% CI 0.90-1.00, P<0.001) trained models all demonstrated a similarly high discriminatory power. An ensemble model, or bootstrap aggregation model, was obtained by grouping all models into one [116] and this model displayed the highest discriminatory power to identify the aortic valve morphology correctly (AUC=0.99, 95% CI 0.98-1.00; P<0.001).

Using a probability cut-off of 0.5, the bootstrap aggregation model demonstrated a sensitivity of 100% and a specificity of 92% to correctly identify BAV, representing a positive predictive value of 92% and a negative predictive value of 100%.

Leaflet calcium volume was higher in patients with BAV ($1066 \pm 652$ vs. $408 \pm 361$ $mm^3$, P<0.0001). Furthermore, leaflet calcium demonstrated a good discriminatory power to correctly classify aortic valve morphology (AUC=0.85, 95% confidence interval [CI] 0.77-0.92, P<0.001).

The training time for one CNN model was about two hours per fold. Average processing time for the trained network was under one second per patient.

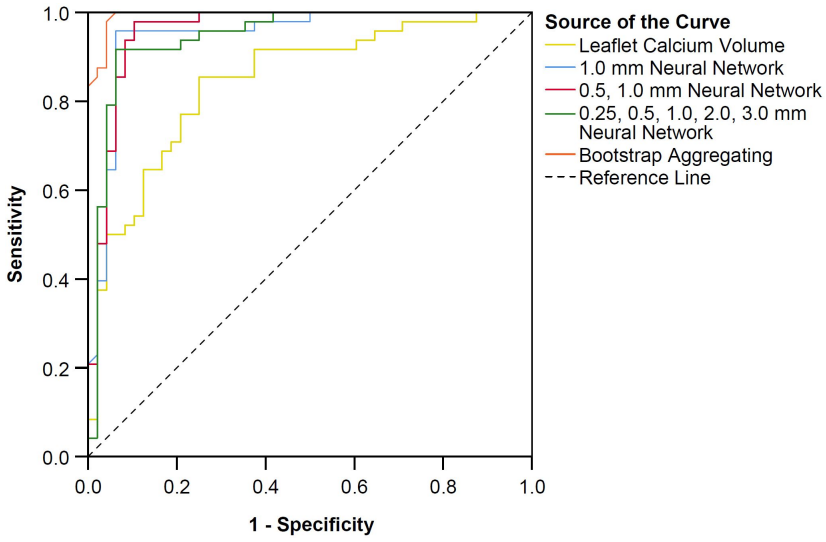



*Figure 4.13: Receiver operating characteristic curve to correctly classify aortic valve morphology.*

The output was captured during the test phase of a trained CNN model to discover what the CNN model learned. The trained kernels developed functions such as edge detection and calcium identification (Figure 4.14).

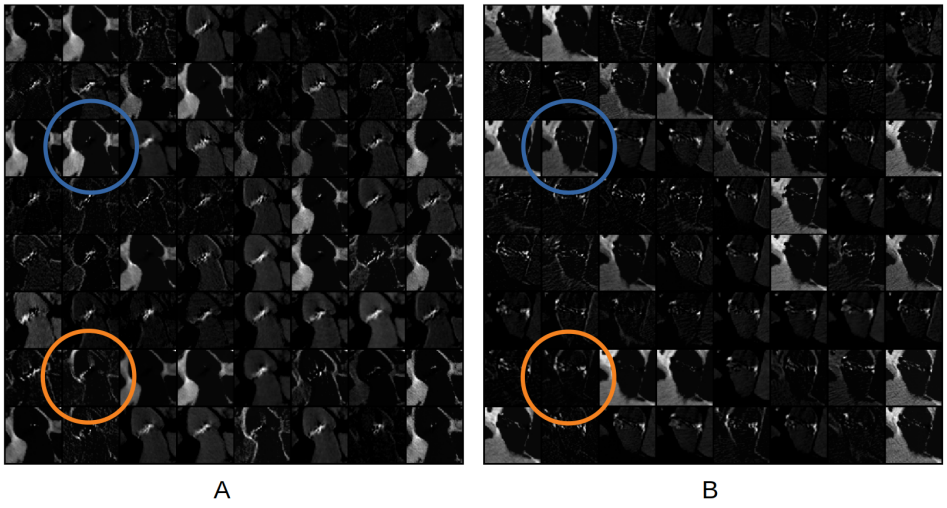



*Figure 4.14: Output of the second 3D convolutional layer of a bicuspid patient (A) and a tricuspid patient (B). The kernels are capable of identifying edges (blue) and calcium (orange).*

### 4.2.4 Discussion

As TAVR continues to expand into younger, lower-risk patient cohorts, the correct identification of aortic valve morphology will be paramount. While outcomes of TAVR in BAV have improved with increased operator experience and newer-generation devices [110, 117, 118], the incidence of paravalvular regurgitation and permanent pacemaker implantation remains higher than with surgery [119, 120]. As such, major society guidelines recommend that surgery should be favoured in patients with BAV [59]. Furthermore, for patients undergoing TAVR, the presence of BAV has significant pre- and peri-procedural implications, such as the need to consider the assessment of the supra-annular structure in order to size the transcatheter heart valve prosthesis [111–113], and the consideration of pre-implantation balloon aortic valvuloplasty.

Also, for patients with aortic dilatation, the presence of BAV has important management implication, with major society guidelines recommending a lower threshold for surgical intervention on the aorta, in the presence of BAV [59].

The identification of BAV can be challenging on transthoracic echocardiography, especially in the late stages of aortic stenosis when the valve becomes heavily calcified and leaflet motion is reduced. MDCT imaging offers superior diagnostic accuracy over transthoracic US when compared to the gold standard of operative findings [121, 122]. However, correct identification of BAV on MDCT imaging can also be demanding, especially without the aid of specialised 3D reconstruction tools. It is therefore likely that many patients with BAV are not correctly identified.

In this study, it was demonstrated that a trained 3D CNN had an excellent ability to classify aortic valve morphology from MDCT imaging correctly, with a high degree of both sensitivity (100%) and specificity (92%). These results are very favourable, especially given the relatively small dataset on which the neural network was trained.

BAV is frequently associated with aortopathy [123] such as aortic dilation (Figure 4.9). In our study, three CNN models were trained, based on increasing volumes of anatomical information. A bootstrap aggregation model displayed the highest discriminatory power to classify aortic valve morphology correctly. It is therefore likely that the trained models have learned to identify the subtle anatomical information contained within the ascending aorta associated with BAV.

BAV is also often related with heavy calcification [124]. These findings were confirmed, and thus, aortic leaflet calcium was a reasonable discriminator for the presence of BAV. However, the trained CNN models demonstrated a superior discriminatory power to identify aortic valve morphology correctly.

Although these results are promising, this study entails a few limitations. First, the MDCT images were all taken from patients with severe aortic stenosis and further validation would be required to assess whether the 3D CNN could correctly classify patients without severe aortic stenosis. Next, the CNN model required image pre-processing to identify the location of the three coronary cusps and could be improved through automation of this process. Thirdly, the dataset was relatively small, necessitating the division of our dataset into four cross-validation folds. The limited size of our dataset also meant that the CNN model could only be trained to

classify aortic valve morphology in a binary manner. Our study could, therefore, be improved by a larger dataset, which could potentially be used to create a 3D CNN to subclassify Sievers Type 0, Type 1 and Type 2 bicuspid aortic valve morphology, by using a softmax as the final activation function. Finally, the dataset was balanced with an equal amount of bicuspid- and tricuspid patients due to the small dataset. This equilibrium does not reflect a realistic setting, where bicuspid patients represent approximately 1% of the general population. This study could be repeated with a dataset which reflects this imbalance in which case the NLL should be weighted to order to counter this imbalance. However, the question remains which of the two strategies (balanced vs imbalanced) would yield the most accurate results.

This study used a 3D CNN model to classify a morphology that can be detected on the level of the aortic root/arc and leaflets. This model was trained on CT images of 96 patients and yielded excellent results. When comparing this work with the available literature, there are only a few studies that used a 3D CNN model to perform classifications [90, 91]. The first study classified the EF from stacked sequences of 2D US images and used the data of 5600 patients. The second study classified the LA from patches of CT images using the data of only 12 patients. None of these studies analysed the entire 3D ROI as input. Adding multiple resolutions in the training or testing phase is a known trick in image analysis, and an example in the presented literature overview of this chapter can be found in Emad *et al.* In this study, different resolutions of the same view were added in the channel dimension of the deep learning model to increase the field of view from aortic leaflet to aortic root/arc.

## 4.3 Conclusion

In the first part of this chapter, a literature overview of cardiac classification with deep learning between 2015 and 2019 was given. The studies were chronologically presented in order to show the evolution of deep learning in the past years: starting from 'simple' CNN models to robust methods for classification.

In the second part of this chapter, a method to classify tricuspid and bicuspid aortic valve patients from MDCT images with a custom 3D CNN model was presented. The trained 3D CNN models demonstrated an excellent ability to correctly classify tricuspid and bicuspid aortic valve morphology from MDCT images. The discriminative power of the CNN models is interesting for the field of TAVR, where younger patients, and thus, more and more BAV patients, will be considered for this minimally invasive treatment.

# 5 Segmentation

*To measure the man,*
*measure his heart.*

Malcom Forbes

Segmentation is the division of something into smaller parts or sections. In image analysis, this is the annotation or selection of one or many specific areas of interest. Segmentation methods and applications can also be found in medical image analysis literature. Given the recent upswing of deep learning, many deep learning segmentation strategies have been proposed in the field of medical image analysis. This chapter starts with a literature overview of cardiac segmentation methods that use deep learning. Afterwards, two novel methods that use segmentation with deep learning are presented.

## 5.1 Literature overview of cardiac segmentation with deep learning

In this section, a literature overview of cardiac segmentation with deep learning is presented. The presentation of this overview is similar to the previous chapter and is also limited to the work related to cardiovascular segmentation with deep learning for brevity and relevance reasons.

### 5.1.1 Before 2016

The authors mentioned in this section proposed pioneering studies where the focus was to provide insights in gradient-based learning. However, the architecture choices

were primitive compared to the available literature; for example, AlexNet was already available in 2012. Also, the sizes of the datasets were small. Table 5.1 contains a short summary of the discussed work.

| Authors | Scope | Year | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|---|
| Carneiro *et al.* [125] | LV segment. with DBN | 2012 | 14 | US | DBN | 2D |
| Carneiro and Nascimento [126] | Tracking LV endocardium | 2013 | 14 | US | DBN | 2D |
| Ngo and Carneiro [127] | LV segment. with level set | 2013 | 45 | MR | DBN | 2D |

*Table 5.1: Summary of cardiac segmentation with deep learning before 2016.*

#### 5.1.1.1 Detailed literature overview

Carneiro *et al.* proposed a new method to segment the left ventricle (LV) from two-dimensional (2D) ultrasound (US) images. The method used two deep belief network (DBN) models: one classifier for object detection, and one classifier for segmentation. The architecture of a DBN model is similar to the architecture of a neural network, but they are trained differently. The first model was trained to recognise the correct view, and the second model was trained to detect the LV contour from 2D ultrasound images. The method was evaluated in 14 patients, and the authors reported high true-positive rates (0.95) and low false-positive rates (0.008) [125].

Going one step further, Carneiro and Nascimento tracked the endocardium of the LV from US images. A deep neural network was used as a motion model and combined systolic and diastolic motion patterns. A mean absolute distance of 0.94 between prediction and ground truth was reported [126].

A novel method was proposed by Ngo and Carneiro, where a pre-processing step and a DBN were combined to segment the LV from a limited number of magnetic resonance (MR) images. The pre-processing step introduced a priori knowledge about the segmented subject, whereas the DBN handles the parameter tuning required to segment the LV. The method was evaluated in 15 patients and achieved an average Dice score of 90%. The authors reported 2-3 minutes per patient, which was comparable to state of the art at that time [127].

### 5.1.2 2016

Different types of segmentation methods can be identified in 2016 [79, 128, 129]. The first method used an encoder followed by a fully-connected layer, which was translated into the segmented 2D image (Figure 5.1). The second method used an encoder and a transposed convolution layer to segment a ROI (Figure 5.2). And the third method, already introduced in chapter 4, used a patch-based method to classify pixels (Figure 4.2). It is fascinating to observe that different types of segmentation methods with different image modalities were used to perform a similar task, which shows the potential of deep learning. Table 5.2 contains a short summary of the discussed work.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Avendi *et al.* [128] | LV segment. with deform. models | 45 | MR | CNN | 2D |
| Rupprecht *et al.* [130] | Deep active contours | 45 | MR | CNN | 2D |
| Tan *et al.* [131] | LV segment. | 45 | MR | CNN | 2D |
| Chen *et al.* [132] | LV segment. | 42894 | US | CNN | 2D |
| Ghesu *et al.* [133] | Marginal space deep learning | 150 | US | FC | 3D |
| Nascimento and Carneiro [134] | Manifold learning with DBN | 14 | US | FC | 2D |
| Zhen *et al.* [135] | Bi-ventricular volume estimation | 100 | MR | CNN | 2D |
| Tran [129] | Bi-ventricular segment. | 45 | MR | CNN | 2D |
| Yang *et al.* [136] | LV myocardium segment. | 33 | MR | CNN | 2D |
| Andermatt *et al.* [137] | Segment. of the brain | 20 | MR | RNN | 3D |
| Dong *et al.* [138] | Direct ESV, EDV estimation with random forest | 30 | US | CNN | 3D |

*Table 5.2: Summary of cardiac segmentation with deep learning in 2016.*

#### 5.1.2.1 Detailed literature overview

Avendi *et al.* proposed a novel method to segment the LV from short-axis 2D MR images with two deep learning models. The first model was a convolutional neural network (CNN) model and was used to determine the square region of interest (ROI) in each slice (Figure 5.1). The second model was a fully connected auto-encoder. This model was trained to determine the contours of the LV from each ROI. The method employed a deformable model to improve the final segmentation and was evaluated in 15 patients. The evaluation yielded an average Dice score of 96.7% with a detection time of 1 second per patient [128].

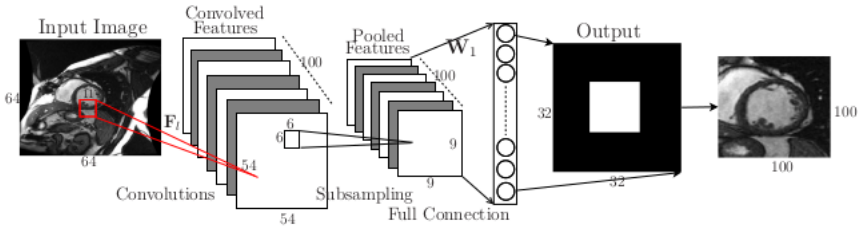



*Figure 5.1: A deep learning architecture to segment the LV from MR images. Image from [128].*

In Rupprecht *et al.*, a contour detection method was proposed. A framework for active contours was used, and a simplified AlexNet model evaluated the proposed contour. Patches from the contour were given to the CNN model which predicted the vector field pointing from the center of each patch to the closest point in the actual contour. This work was evaluated on medical and non-medical images [130].

A new method was proposed by Tan *et al.*, where the LV endocardium was segmented from MR images with two CNN models. The first model was trained to detect the center of the LV endocardium from 2D MR images. The second CNN model was trained to detect the endocardial contour from a 2D slice in polar coordinates. The method was validated and yielded an average Dice score of 88% [131].

A large study was performed by Chen *et al.*, where the data of 42894 patients were used. The method segmented the LV from 2D US images with two CNN models. The first model segmented the LV mask from a single 2D image (from multiple views) and the second model was used to refine the segmentation by patch-wise contour segmentation. The models were trained to minimise the negative log-likelihood (NLL) and regularised with L2 regularisation. The method was evaluated on images of 8533 patients, which yielded an average Dice score of 89.2% [132].

Ghesu *et al.* proposed a new method to segment the aortic valve from US images. The method first detected a three-dimensional (3D) bounding box by using three deep neural networks to classify the correct candidate patches. After the 3D bounding box detection, a second part of the proposed method segmented the aortic valve from US images after which the predicted mesh was generated. The validation yielded an average distance to the ground truth mesh of 0.9 mm [133].

A new approach was proposed by Nascimento and Carneiro, to detect and segment the LV from US. The method used manifold learning, which partitions the data into several patches and learns to segment these patches. A DBN classified these patches in order to fuse them into a final segmentation [134].

In Zhen *et al.*, a method to estimate the volumes of the two ventricles from MR was proposed. The method employed a CNN model as an image encoder and trained a regression forest with the CNN features. The authors used leave-one-out cross-validation on a dataset of 100 patients to validate the method [135].

Tran also proposed bi-ventricle segmentation. The fully convolutional neural network (FCNN) was trained and validated on a dataset containing the MR images of 45 patients. The validation yielded a Dice score of 96% and 92% for the segmentation of the epi- and endocardium [129]. Figure 5.2 depicts a schematic overview of the architecture that was used in this study.

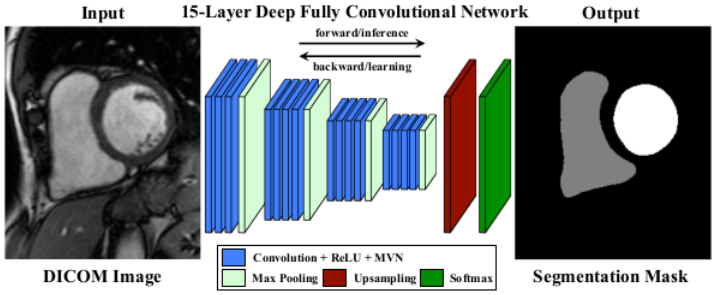



*Figure 5.2: Schematic overview of a segmentation architecture in 2016: 15-layer deep FCNN. Image from [129].*

A new method to segment the myocardium of the left ventricle was proposed by Yang *et al.* A CNN model was used to segment the myocardium from MR images and consisted of six convolutional layers, two pooling layers and one transposed convolutional layer. The validation yielded an average Dice score of 75% [136].

Andermatt *et al.* proposed a new method to segment MR images of the brain with a recurrent neural network (RNN). This work is only included to identify all possible deep learning methods for segmenting medical images [137].

In Dong *et al.*, a method to directly predict the end-diastolic volume (EDV), end-systolic volume (ESV) and ejection fraction (EF) from 3D US images was proposed. The feature representations of CNN models were used to train random forests. The predicted EDV, ESV and EF were compared to the ground truth and yielded $R^2$ values of 0.85, 0.87 and 0.86 [138].

### 5.1.3 2017

The studies in this section show that segmenting medical images with deep learning becomes more robust. The presented studies explore the usage of deep learning elements such as optimisers, batch normalisation, residual connections, model architectures and cost functions [139–141]. The dataset sizes are relatively similar when comparing to the dataset sizes in 2016. Table 5.3 contains a short summary of the discussed work.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Giannakidis *et al.* [142] | Two-path RV segment | 26 | MR | CNN | 3D |
| Li *et al.* [143] | Whole heart segment | 20 | MR | CNN | 3D |
| Luo *et al.* [144] | Direct ESV and EDV estimation | 937 | MR | CNN | 2D |
| Luo *et al.* [145] | localisation and segment of RV | 48 | MR | CNN | 2D |
| Yu *et al.* [146] | Fractal net | 20 | MR | CNN | 2D |
| Dou *et al.* [147] | Full heart segment | 20 | MR | CNN | 3D |
| Wolterink *et al.* [148] | Dilated CNN for heart segment | 20 | MR | CNN | 2D |
| Lieman-Sifry *et al.* [139] | LV, RC segment (model eval) | 1143 | MR | CNN | 2D |
| Curiale *et al.* [141] | Myocardial segment (model eval) | 45 | MR | CNN | 2D |
| Tan *et al.* [149] | localise-and-segment method (polar) for LV | 200 | MR | CNN | 2D |
| Liao *et al.* [150] | LV ESV, EDV estimation | 1140 | MR | CNN | 2D |
| Yang *et al.* [151] | localise-and-segment method for LV | 45 | MR | CNN | 2D |
| Romaguera *et al.* [140] | Endocardium segment from LV (optim eval) | 45 | MR | CNN | 2D |
| Avendi *et al.* [152] | localise-and-segment method for RV | 48 | MR | CNN | 2D |
| Ngo *et al.* [153] | DL & level set segment epi- and endocardium | 45 | MR | FC | 2D |
| Kabani and El-Sakka [154] | LV localisation | 25 | MR | CNN | 2D |
| Yang *et al.* [155] | Multi-region segment (loss eval) | 60 | CT, MR | CNN | 3D |
| Baumgartner *et al.* [156] | RV, LV segment (loss and network eval) | 100 | MR | CNN | 2D |

*Table 5.3: Summary of cardiac segmentation with deep learning in 2017.*

#### 5.1.3.1 Detailed literature overview

Giannakidis *et al.* proposed a novel strategy to segment the right ventricle (RV) using a CNN model with two input paths. The model was trained to classify the voxels of the RV from MR images (similar to [79]). In order to learn the surrounding and detailed anatomy, image patches around the voxels were selected in two resolutions: low and normal. The method was validated on 26 patients and yielded an average Dice score of 80%. Figure 5.3 shows a schematic overview of the architecture and method that was used in this study [142].

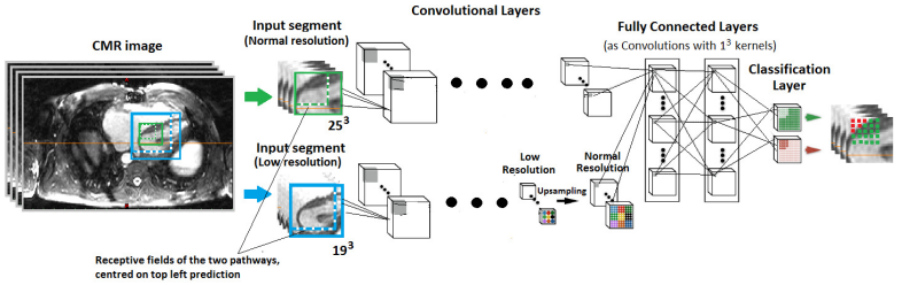



*Figure 5.3: Segmentation through low- and normal resolution patches. The CNN model with two input paths learns the surrounding and detailed anatomy around each voxel. Image from [142].*

In Li *et al.*, a 3D fully convolutional neural network (FCNN) model was used to segment the whole heart from MR images. The model was equipped with dilated convolutional layers. The reported test average Dice score was 70%, which is low when comparing to other segmentation methods [143].

A novel method to predict ESV and EDV from MR images was proposed by Luo *et al.* The CNN model was trained to predict the ESV and EDV from multiple views: base-, mid- and apical slice from the four- and two chamber view of both target phases. The predicted ESV, EDV and EF yielded correlation coefficients 0.92, 0.95 and 0.9 when compared with the ground truth [144].

Luo *et al.* proposed a new strategy to localise and segment the RV from MR images using two CNN models. The first CNN model was used to predict the center of the ROI directly and was optimised by minimising the Euclidean distance between the predicted- and the ground truth center. The second model was trained to segment the ROI. The validation of the localisation was not reported in their results. The segmented endo- and epicardium was validated and yielded average Dice scores of 86% and 84% when compared to the ground truth [145].

The method proposed by Yu *et al.* segmented the whole heart from MR images. A fractal rule constructed the CNN network. The method was validated on ten patients and yielded an average Dice score of 78% on the segmentation of the myocardium, which is low when comparing to other segmentation methods [146].

A method to segment the liver and the whole heart from 3D medical images was proposed by Dou *et al.* The liver was segmented from computed tomography (CT) images and the whole heart from MR images. The authors trained the models with the NLL as a cost function and added L2 regularisation. The authors used many metrics for comparing the predicted segmentation with the ground truth such as volumetric overlap error, volume difference, average symmetric surface distance, and others. The validation yielded an average Dice score of 73% on the segmentation of the myocardium of the whole heart. Figure 5.4 shows a schematic overview of the architecture that was used in this study [147].

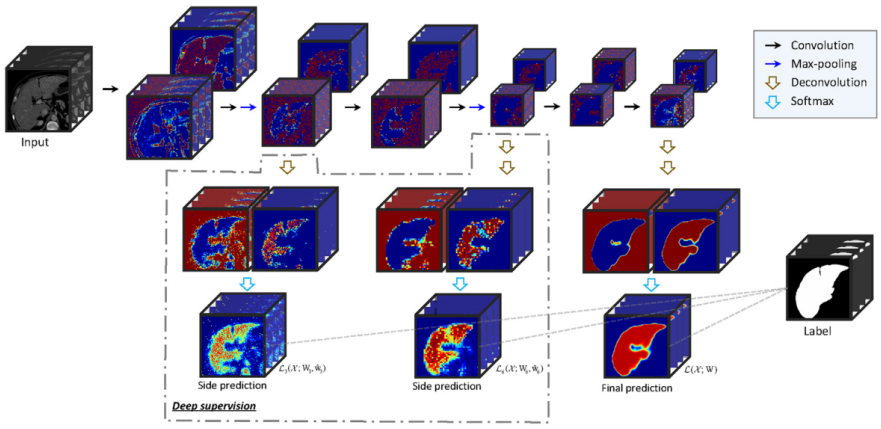



*Figure 5.4: Image flow from 3D medical image to 3D segmentation in 2017. Image from [147].*

Wolterink *et al.* presented a method that used dilated convolutions in the CNN model, which was used to segment the whole heart from MR images. It was shown that dilated convolutions reduced the number of required parameters without limiting the model's field of view. The method was validated on a small test set of 10 patients and yielded an average Dice score of 80% for segmenting the myocardium [148].

Lieman-Sifry *et al.* performed a comparative study between two CNN models. The models were trained to segment the LV and RV from 2D MR images. U-Net inspired the first model, and the second model was inspired by the Inception model and was optimised to reduce the memory- and computation footprint. The authors reported no significant difference between the two models and used relative absolute volume error as metric [139].

Another comparative study was presented by Curiale *et al.* In this work, a method to segment the myocardium of the LV from MR images was proposed. The target model was a 2D CNN model inspired by U-Net. The authors compared the usage of residual connections, batch normalisation and different cost functions, i.e. the Jaccard index and the Dice score. The Jaccard index identifies the similarity between finite sample sets. It was concluded that the model that used batch normalisation, residual connections and was trained with the Jaccard index yielded the highest average Dice score of 90% [141].

Building on top of their work from 2016, Tan *et al.* proposed a two-step method to segment the contours of the LV. The first step localised the center of the LV using a CNN model and the second step segmented polar images which were derived from the center of the LV. The output of the first CNN model was a 2D coordinate and the output of the second CNN model were the epi- and endocardial distances from the center of the LV. The second model used a coarse- and fine-grained version of the input to include more anatomical information. The validation was performed on 100 patients and yielded an average Jaccard index of 77% [149].

In Liao *et al.*, a LV volume estimation technique from MR images was proposed. First, a ROI was selected and next the LV was segmented from this ROI. The final LV volume was estimated by summing all predicted areas from a 2D MR image. The proposed work was 4th at the Second National Data Science Bowl competition of Kaggle in 2015 [150].

Similarly, a method to localise and segment the LV from MR images was proposed by Yang *et al.* The localisation method used a 2D CNN model to predict the top-left and bottom-right corner and the segmentation method used a 2D CNN model (inspired by U-Net). The segmentation model was optimised using the Dice score as the cost function. The method was trained and validated on two datasets containing images of 33 and 45 patients. The validation was performed on 15% of the datasets and yielded an average Dice score of 90% for both datasets [151].

The LV was segmented from MR images in a comparative study between two optimisers by Romaguera *et al.* The CNN model was a 16-layered model that used an encoder and one transposed convolution layer to predict the segmentation mask. The authors compared the SGD and the RMSProp optimiser, each with a learning rate decay strategy (multi-step for SGD and inverse for RMSProp). The method was cross-validated with 5-fold cross-validation, and an average Dice score of 92% and 90% was obtained for the SGD and the RMSProp [140].

Avendi *et al.* proposed a method to segment the RV from MR images using two 2D CNN models. The method used a localisation- and a segmentation model. The localisation model was a regression model that was trained to predict the x- and y coordinates of the center of the RV in the MR slice. This center was used to select the ROI which was segmented by the second CNN model. The method was validated with 32 patients and yielded an average Dice score of 82,5% [152].

A novel method was proposed by Ngo *et al.*, where the endo- and epicardium of the LV was segmented from MR images using a combination of deep learning and image analysis methods. The method is subdivided in several steps starting with a 2D ROI mask detection from MR slices using a DBN. From the ROI, the endocardium was segmented using another DBN, level set contour approximation and by using a shape prior. A similar approach was performed for the epicardium. The method was validated on 15 patients and yielded an average Dice score of 86% and 92% for the segmentation of the endo- and epicardium [153].

In Kabani and El-Sakka, a method to localise and segment the LV at end-systolic and end-diastolic phase from MR images was proposed. The localisation method was similar to [78] and the segmentation method used the U-Net model. After the segmentation, the ESV, EDV and EF were derived and validated [154].

Yang *et al.* proposed a method to segment seven partitions of the heart from CT and MR images. U-Net inspired the 3D CNN model, and the authors evaluated different cost functions. The weighted NLL was used as a baseline and compared with a hybrid cost function that combined the weighted NLL and the weighted Dice score. The method was trained and validated on a CT and a MR dataset each containing images of 60 patients. The validation yielded an average Dice score of 75% and 77% for the NLL and the hybrid cost function [155].

The study of Baumgartner *et al.* provided an overview of different methods to segment the LV, RV and myocardium from MR images using CNN models. The authors compared three 2D and one 3D architecture(s): VGG (with upsampling), U-Net, a modified U-Net (with a smaller upsampling path) and 3D U-Net [157]. The authors also compared three different cost functions: the NLL, the weighted NLL and the Dice score [25]. The experiments were conducted on a dataset of 100 patients (with an 80:20 training-test split). In almost all of the experiments, the 2D U-Net or the modified version of it, yielded the highest average Dice score. The weighted NLL yielded the highest average Dice score when training the U-Net model [156]. No explanation was given why the V-Net architecture [25] was not used (although the cost function from the same study was used).

### 5.1.4 2018

Segmenting medical images with deep learning was almost standardised in 2018 when looking at the available literature. Some studies presented interesting work while also evaluating deep learning elements such as optimisers [158]. There are more studies which use a 3D CNN model and more studies that analyse sequences of 2D images. The datasets are larger when compared to the studies in 2017. Table 5.4 contains a short summary of the discussed work.

| **Authors** | **Scope** | **Patients** | **Modality** | **Method** | **Dim.** |
|---|---|---|---|---|---|
| Payer *et al.* [159] | Whole heart segment with two models | 20 | CT | CNN | 2D |
| Xue *et al.* [95] | LV quantification | 145 | MR | RNN | 2D+t |
| Dong *et al.* [160] | LV segment | 45 | US | CNN | 2D |
| Oktay *et al.* [161] | LV segment with super resolution | 45 | CT, US | CNN | 2D |
| Zhang *et al.* [162] | RV segment with multi-task network | 514 | MR | CNN | 2D |
| Romaguera *et al.* [158] | Myocardial segment (optimiser eval) | 45 | MR | CNN | 2D |
| Tan *et al.* [163] | localise-and-segment method (polar) for LV | 1340 | MR | CNN | 2D |
| Du *et al.* [164] | Bi-ventricle contour detection | 145 | MR | CNN | 2D |
| Zheng *et al.* [165] | localise and segment ventricles | 3834 | MR | CNN | 2D |
| Savioli *et al.* [166] | Temporal LV segment | 68 | MR | RNN | 2D+t |
| Isensee *et al.* [167] | LV and RV cavity and LV myocardium segment | 150 | MR | CNN | 3D |
| Vigneault *et al.* [168] | Multi-regional cardiac segment | 63 | MR | CNN | 2D |
| Duan *et al.* [169] | localise and segment ventricles | 2480 | MR | CNN | 2D |
| Bai *et al.* [170] | Segment ventricles | 4875 | MR | CNN | 2D |
| Mortazi *et al.* [171] | Whole heart segmentation | 60 | CT, MR | CNN | 2D |
| Zotti *et al.* [172] | LV, RV and myocardium segment | 100 | MR | CNN | 2D |
| Patravali *et al.* [173] | LV, RV and myocardium segment | 150 | MR | CNN | 3D |
| Winther *et al.* [174] | LV and RV segment | 1031 | MR | CNN | 2D |

*Table 5.4: Summary of cardiac segmentation with deep learning in 2018*

#### 5.1.4.1 Detailed literature overview

In Payer *et al.*, a method to localise and segment the whole heart from CT and MR images was proposed. The method employed two models, which were both inspired by U-Net. The first CNN model applied heatmap regression to detect the center of the heart. This model was trained with a low-resolution down-sampled version of the medical images. Once the center of the heart was detected, the second CNN model was trained to segment the whole heart into seven different segments. The average Dice score was 79% for the MR image dataset and 88% for the CT dataset [159].

A novel method to quantify the LV from sequences of MR images was proposed by Xue *et al.* The method used a CNN encoder and two recurrent models. A CNN model encoded the sequences of images into vectors (of length 1000). These sequences of vectors were used in the recurrent models. The first RNN was trained to estimate the indices (such as LV area, wall thicknesses, and others) and the second RNN was trained to estimate the cardiac phase [95].

Dong *et al.* proposed a method to segment the LV from 2D US in two steps. The first step used a trained 2D CNN model to segment the US slices into segmentation masks, which resulted in a coarse segmentation. This segmentation was corrected in a second step. During this step, the segmented masks were re-centered using the radius and centers of the masks. The validation yielded an average Dice score of 91% and 88% of the ED and ES volumes [160].

A comparative study between optimisers was presented by Romaguera *et al.* where the myocardium of the LV was segmented from MR images. The model used a known approach: starting with a CNN encoder, which was followed by a transposed convolutional layer and concluded by a softmax layer (Figure 5.5). The validation of the method was performed on 45 patients and yielded an overall average Dice score of 90%. The authors also performed a thorough study on the optimisers, where different optimisers were compared with each other on the same task. The included optimisers yielded the following average Dice score: RMSProp (93%), SGD (92%), Nesterov (90%), Adam (89%), Adadelta (88%) and Adagrad (88%) [158].

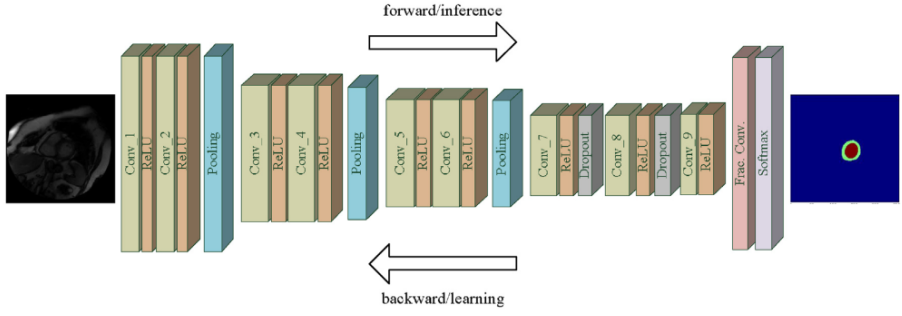



*Figure 5.5: An deep learning architecture to segment medical images in 2018. Image from [158].*

Going a step further, the method proposed by Oktay *et al.* segmented anatomic features from US and MR. The method included super-resolution, which is a technique to enrich the input images with more information. The segmentation method was validated on the US image data from 30 patients and yielded an average Dice score of 91%, which is high, considering the coarse image quality of US images [161].

In Zhang *et al.*, the segmentation of the RV from MR images was proposed using a CNN model with three different tasks. The first task was to classify the size of the contour of the RV, the second task consisted of segmenting the entire image, and the third task was responsible for segmenting the smaller, cropped image. The method was validated on 30% of the dataset and yielded an average Dice score of 80% [162].

Building on top of their previous work [131, 149], Tan *et al.* used two cohorts of 200 and 1140 patients to train their models [163].

Du *et al.* proposed a method to determine the bi-ventricle contours from MR images. The method used a DBN to encode the input image into features. From these features, 100 points were proposed as a contour. The method was validated on 145 patients and yielded an average Dice score of 91% and 94% when segmenting the endo- and epicardium [164].

In Zheng *et al.*, ventricle segmentation was performed in two-steps. First, the ROI was determined by a U-Net model. The model was trained to classify a region including both ventricles. Secondly, an adaptation of U-Net was used to perform the segmentation. The model used the MR slices and the previously segmented mask as input. The output of the segmentation model was a segmented mask of the current MR slice and was used as input for the next MR slice (Figure 5.6). This method used 3078 patients for training and 902 patients for validation. The patients were selected from four different datasets. The validation yielded average Dice scores from 71% to 88% [165].

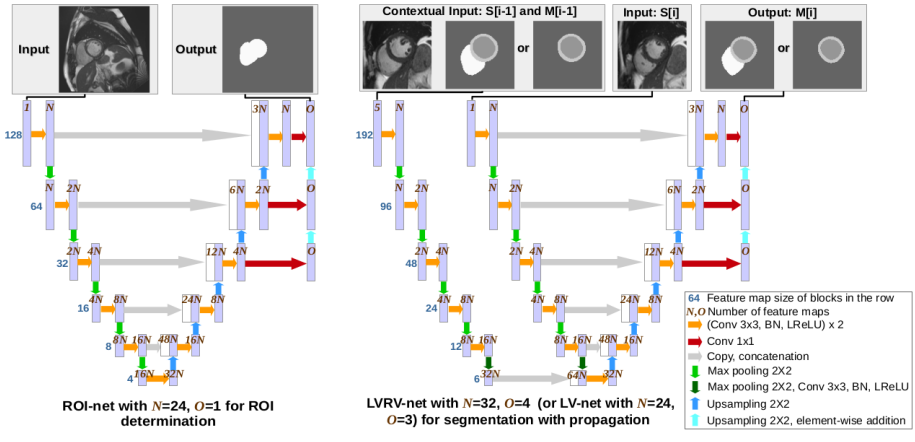



*Figure 5.6: Bi-ventricle segmentation using a dual input CNN model. Image from [165].*

A novel recurrent model was proposed by Savioli *et al.*, where the LV was segmented from temporal 2D MR images. The model was inspired by U-Net and used a gated recurrent unit [175] in between the encoding and decoding path. The method used a cohort of 68 patients, and the validation was performed on 15% of the patients, which yielded an average Dice score of 98% [166].

The novel segmentation method proposed by Isensee *et al.*, where the RV cavity, LV cavity and LV myocardium were segmented from MR images. An ensemble of 2D and 3D U-Net models was used for the segmentation. The method used 150 patients from which 50 were used for validation, which yielded an average Dice score of 90%. The trained models were used as feature generators to train classifiers to distinguish healthy from unhealthy patients [167].

The previous method was used as a point of reference in Vigneault *et al.*, where the authors proposed a novel architecture. The architecture was trained to segment five regions of cardiac MR image. The validation yielded an average Dice score of 89% when segmenting the myocardium, which was lower than the state of the art. Only the segmentation of the blood pools yielded a higher average Dice score [168].

In Duan *et al.*, a method to segment the LV and RV from 2D MR images was proposed. The method used a CNN model that was trained to segment the ROIs and locate their center. Two models were trained: one with high- and one with low-resolution images. The method was validated on 600 patients and yielded average Dice scores between 69% and 95% [169].

Similarly, Bai *et al.* proposed a method to segment the LV and RV from MR images. The method used a 2D FCNN model which was trained and validated on a large dataset of 4875 patients. The validation itself was performed on 600 patients and yielded an average Dice score of 90% [170].

A novel method to segment the whole heart from MR and CT images was proposed by Mortazi *et al.* This method used three CNN models, which were inspired by U-Net and were trained with three orthogonal views: axial, sagittal and coronal. After the segmentation step, a post-processing step fused all predictions. The validation of the whole heart segmentation yielded an average Dice score of 85% and 90% for MR and CT [171].

In Zotti *et al.*, a method to segment the LV, RV and myocardium from MR images was proposed. The model resembled U-Net with an encoding and decoding path, where the encoding path was responsible for reducing the spatial resolution. The authors also employed a custom cost function, which included segmentation, contour and point detection. The method was validated on 25 patients and yielded an average Dice score of 90% [172].

Going a step further, a 3D CNN model was used to segment the LV, RV and myocardium from MR images in Patravali *et al.* The method was validated on 20 patients which yielded an average Dice score of 90% [173].

Winther *et al.* proposed a method to segment the LV and RV from MR images. The predicted segmentations were used to estimate the ESV, EDV and other cardiac parameters. The authors used a CNN model similar to U-Net and the method was validated on four different datasets, which yielded an average Dice score of 88% [174].

### 5.1.5 2019

The number of available literature at the moment of writing was less than in 2018. Table 5.5 contains a short summary of the discussed work.

| Authors | Scope | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|
| Khened *et al.* [176] | LV and RV segmentation | 1140 | MR | CNN | 2D |
| Curiale *et al.* [177] | LV function prediction | 140 | MR | CNN | 2D |

*Table 5.5: Summary of cardiac segmentation with deep learning in 2019*

#### 5.1.5.1 Detailed literature overview

In Khened *et al.*, a method to segment the LV and RV from MR images was proposed. The authors proposed a 2D CNN model that was inspired by DenseNet and compared it with other networks. The validation was performed on three public datasets and yielded an average Dice score of 90% [176].

A novel method to quantify the LV was proposed by Curiale *et al.* The quantification included the LV mass, ejection fraction and stroke volume which were derived from MR images. The U-Net model was used as a baseline but was expanded with residual blocks. The models were trained to detect the myocardial tissues, and the validation yielded an average Dice score of 86% [177].

### 5.1.6 Meta-analysis of the literature overview

The above-mentioned literature can be categorised into the following categories; anatomical region, pre-processing step and reported metric. Since this literature overview was limited to the cardiovascular domain and deep learning, the available anatomical regions are limited. However, when matching other aspects such as technical differences and reported metrics, certain perspectives can be observed and will be discussed in this section.

#### 5.1.6.1 Anatomical region

In the literature overview, the following anatomical regions are discussed: the left ventricle, the right ventricle, both ventricles and the whole heart. The subdivision of endo-, epi- and myocardium is excluded for brevity reasons.

**Left ventricle** The following studies introduce novel methods to analyse medical images and segment the left ventricle: [95, 125–128, 130–132, 134, 136, 140, 141, 149–151, 153, 154, 158, 160, 161, 166, 177].

From these studies (and where the information was available), the evaluation metric was extracted. The highest Dice score of 98% was found in the study of Savioli *et al.* [166], and the lowest Jaccard index of 77% was found in Tan *et al.* [149] (Jaccard index and Dice score are similar metrics). The first study used a deep learning model based on U-Net with recurrent layers and used the MR images of 68 patients (15 for validation). The second study used a CNN model and the MR images of 200 patients

(100 for validation). Based on these results, one may develop a preference for the study with the highest validation value; however, when looking at the ground truth data, the following can be observed. In the study with the highest validation value, the ground truth displayed the highest level of detail and followed the ventricular wall (Figure 5.7 A). This level of granularity was not present the study with the lowest validation value, where the ground truth consisted of almost elliptical shapes coarsely following the ventricular wall (Figure 5.7 B). This difference between the two studies could explain the vast difference between the validation values. If the segmentation ground truth follows the delineations in the anatomy, the deep learning models display a higher accuracy. However, it remains to be investigated if there are other reasons, such as superior architectural design, which can explain the difference between the two studies.

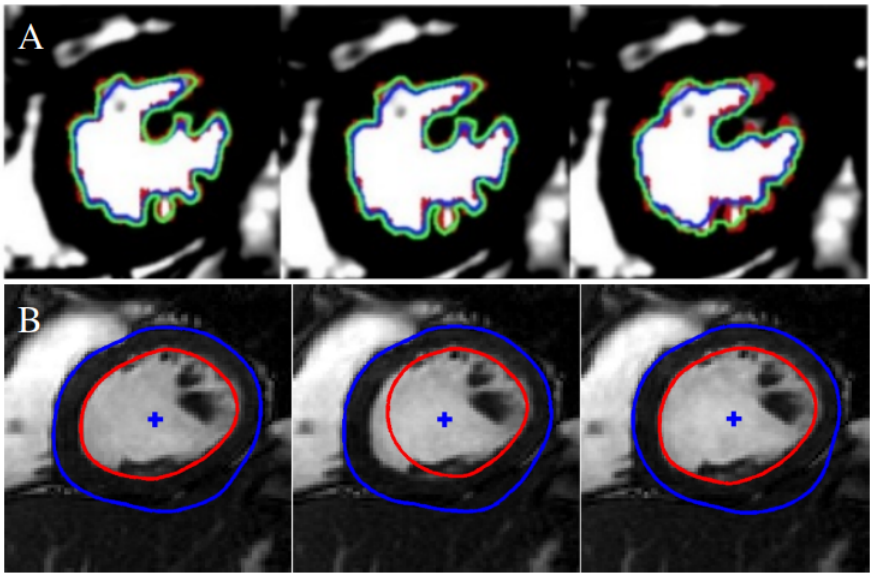



*Figure 5.7: Comparison of two methods that segmented the left ventricle from MR. At the top(A), a sample of the validation data from Savioli* et al. *[166] is shown. The blue and green line indicates the output of a model and the red line indicates the clinical ground truth. At the bottom (B), a sample of the ground truth validation data from Tan* et al. *[149] is shown. The blue line indicates the epicardium and the red line indicates the endocardium.*

**Right ventricle** The following studies introduce novel methods to analyse medical images and segment the right ventricle: [142, 145, 152, 162].

These studies used the Dice score as a validation metric, which was between 80% and 86%. Two studies used the same dataset [178], which contained the MR images of 48 patients and the ground truth segmentation of the epi- and endocardium of the right ventricle. These two studies are compared in the next section.

The first study obtained an average Dice score of 85% using 32 patients for validation (Luo *et al.* [145]). The second study reported an average Dice score of 82.5% and also used 32 patients in their validation (Avendi *et al.* [152]). Both studies included a pre-processing step to detect the region of interest. Moreover, the first study reported a worse performance if this pre-processing step was excluded. Both studies used a similar pre-processing step where a CNN regression model predicted the center of the region of interest. However, in the segmentation step of both studies, differences can be observed. The study with the slightly higher validation Dice score used a rudimentary model with only 5 CNN layers to segment the right ventricle, whereas the other study used a fully connected regression model to predict the contours directly. After this contour detection, another step was required to obtain the final segmentation. Since the validation values were similar, one may conclude that both approaches are equivalent to segment the right ventricle from MR images; however this should be verified.

**Both ventricles** The studies in the two sections above were limited to a single ventricle and thus analysed a smaller region. The following studies introduce novel methods to analyse medical images and segment both ventricles: [129, 135, 139, 156, 165, 167, 169, 170, 172–174, 176]. Segmenting two ventricles implies that the models will need to learn to segment a larger region of interest which can be challenging.

All of the aforementioned studies (except two [135, 139]) use the Dice score as a validation metric which ranged from 88% to 96%. There are six studies which use a patient cohort greater than 1000 patients [139, 165, 169, 170, 174, 176].

Since the anatomical region includes both ventricles, different methods are used to handle the image-space. Some studies use a low-and high-resolution version of the medical images [169] whereas others use a method to detect the region of interest first [164, 165, 169] in order to limit the amount of image-space.

**Whole heart** In this last section, studies that segmented the whole heart as the anatomical region are discussed [143, 146–148, 155, 159, 168, 171]. The size of the image volume becomes challenging when segmenting the whole heart.

Two studies are compared, which obtained a similar validation Dice score and used an equivalent amount of patients. The first study used an hourglass-shaped CNN model to refine the output of the first U-Net model [168]. The second study used a CNN model which was inspired by U-Net [171]. Both methods use 2D input images for the models and fixed size and thus crop of pad the medical images in order to overcome the variational size of medical images.

#### 5.1.6.2 Pre-processing step

Some methods used a pre-processing step in order to detect the center of bounding box the region of interest [128, 131, 133, 145, 149–153, 159, 165, 169, 179]. These studies used a method (albeit with deep learning) to detect the region of interest from the medical images prior to the actual segmentation.

Some of the novel methods to segment the left ventricle used the same dataset which contained the MR images of 45 patients [127, 128, 130, 131, 140, 141, 151, 153, 158]. The study with the highest reported validation Dice score of 96.7% [128] included a pre-processing step. This may indicate that adding a pre-processing step before segmenting increases the accuracy of the model.

#### 5.1.6.3 Reported metric

Finally, the reported metrics (and the results thereof) can also be compared. It is not possible to compare methods with different anatomical regions, image modality, image source, ground truth or cohort size by a single metric. However, when observing the reported validation values of the anatomical region categories, the following can be observed. The studies which analysed the left ventricle (21) displayed a median Dice score of 90% whereas the studies which segmented the right ventricle (4) displayed a median Dice score of 80%. The studies which analysed both ventricles (12) obtained a median Dice score of 90%, and the studies of the segmentation of the whole heart (8) obtained the lowest median Dice score of 79%. This difference may indicate that segmenting the whole heart is the most challenging task, followed by segmenting the right ventricle. It remains to be established if methods that segment the left ventricle are also applicable to segment the right ventricle. Also, it remains to verified if deep learning models are better at a single (simple) segmentation task than multiple segmentation tasks at once (left ventricle vs whole heart).

### 5.1.7 Conclusion

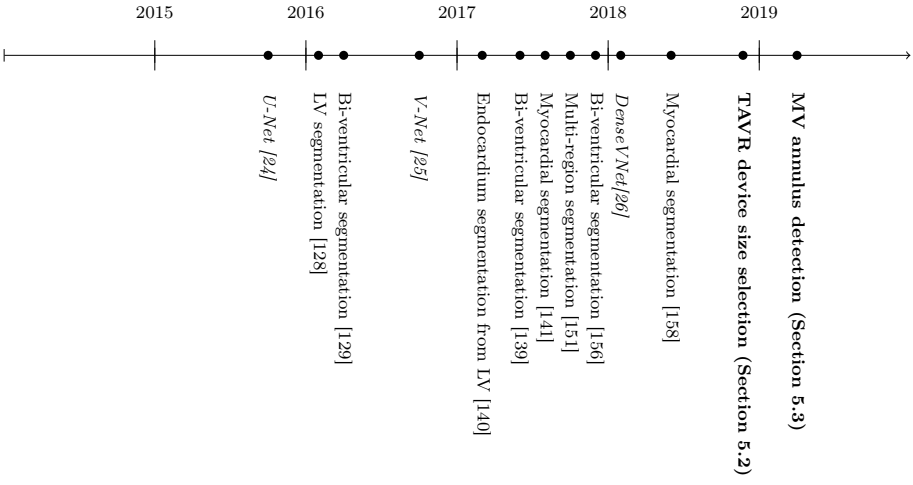



*Figure 5.8: Chronological overview of the most significant milestones of the presented literature, the TAVR device size selection study and the mitral valve annulus detection study.*

This concludes the literature overview of cardiac segmentation with deep learning. As in the previous chapter, the evolution of deep learning can also be noticed in this overview, starting from pioneering studies which were applied to just a few patients [125, 127] towards a standardised method for segmenting medical images in 2018. In the next sections, novel methods that apply deep learning to segment medical images are presented. A timeline with the most significant milestones in the presented literature overview and these novel methods are depicted in Figure 5.8.

## 5.2 Enabling automated device size selection for TAVR

### 5.2.1 Introduction

Before a TAVR procedure, the patient's cardiovascular anatomy - including the aortic valve, aortic root/arch and possible access vessels, have to be assessed in detail. Since TAVR is a closed-chest procedure, the pre-operative planning is image-based. The MDCT images are analysed by a trained operator to obtain the dimensions of the aortic annulus, amount of calcium, and other measurements in order to assess risk factors and select a correct prosthesis size. Correct sizing is of paramount importance because over-sizing may cause aortic annulus rupture or conduction abnormalities [180] and under-sizing may cause device migration or valvular regurgitation. Therefore, image quality and operator experience are essential to select correct device size selection and influence the optimal procedural outcome.

The number of TAVR procedures is increasing each year [181], and as a result, scalability of the complete procedure, including pre-operative planning becomes an important aspect. To address this, experienced operators can enlarge their volume of TAVR cases, for example, by increasing the efficiency. On the other hand, many new operators will need to be trained, which logically leads to increased risks due to their limited experience. When focusing on the pre-operative planning, accurate automated detection of the aortic annulus dimensions directly from MDCT images could not only increase efficiency but at the same time reduce operator variability, thereby minimising the impact of experience on TAVR sizing.

In this study, a deep learning method is presented that can predict aortic annulus perimeter and area automatically. The method is validated against an inter-observer variability study to assess its accuracy. As a final step, the impact of the proposed method on the prosthesis size selection for both the Edwards Lifesciences and Medtronic transcatheter aortic bioprostheses was evaluated.

### 5.2.2 Materials and methods

#### 5.2.2.1 MDCT imaging

This retrospective study used the anonymised data of 473 patients collected from multiple centra. The mean age of this cohort was 80.8 $\pm$ 7.2 years, and 55% of the patients were female. There were 36 bicuspid patients in this cohort. The patient

data consisted of volumetric MDCT images, which were acquired to support the pre-operative phase of a TAVR procedure. Therefore, all MDCT images were contrast-enhanced and contained a certain degree of aortic stenosis. The average row, column, slice thickness and pixel-spacing of the MDCT images were 512.1, 511.7, 0.8 and 0.5 mm.

The aortic annular plane (AAP) was manually identified by an expert from the volumetric MDCT images using the standard method [182] and validated by a second observer.

#### 5.2.2.2 Manual detection

The aortic annulus was segmented from the aortic annular planes for all patients by observer one. The data of the first observer was considered the ground truth in this study. A second observer re-segmented the aortic annulus from the AAP of 118 randomly selected patients in order to assess the inter-observer variability. Both observers applied the same manual method, which consists of visual detection of the aortic annulus within the AAP and annotating it using Materialise Mimics Innovation Suite 18 (Mimics, Leuven, Belgium). The manual segmentation of the aortic annulus was considered a robust yet time-consuming procedure.

#### 5.2.2.3 Automatic detection

This study was focused on automating manual segmentation and deriving clinical patient-specific information. The pre-processing of the ground truth images and aortic annulus annotations was necessary in order to prepare the data for training of the deep learning models.

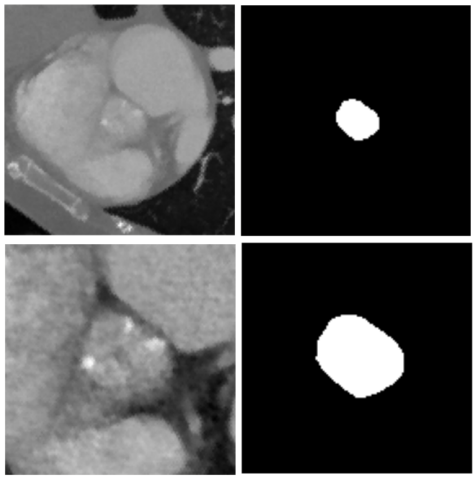


*Figure 5.9: Example of the aortic annular planes and the accompanying binary masks. The resampled and clipped aortic annular planes (left) and the binary masks (right) with different resolutions 1.0 mm (top) and 0.5 mm (bottom).*

**Pre-processing** The aortic annular planes were clipped and resampled in order to fit the model's input. The aortic annular planes were resampled to an isotropic resolution of 1 mm to create a homogeneous dataset. The original Hounsfield units were preserved during the resampling process, which was performed with cubic spline interpolation. Since the average slice thickness and pixel-spacing were 0.8 and 0.5 mm in this dataset, the level of detail was reduced by resampling all images to a 1 mm resolution. In order to prevent this data loss and increase the level of detail, the second resolution of 0.5 mm was also exported, which is a known trick in literature [142, 159, 169]. As the deep learning network expected a 128 x 128-pixel plane as input, the resampled aortic annular planes were clipped around the center of the aortic annulus. The time to pre-process the data of a single patient was within seconds.

The ground truth annotations were used to create binary masks that would mark the aortic annulus within the aortic annular plane for both resolutions (Figure 5.9).

**Architecture** The baseline architecture in this study was U-Net [24]; however, it was observed that this architecture was not able to yield the required details. It is therefore that this study used a deep learning architecture inspired by U-Net and ResNet [52], and consisted of a down-scaling- and an up-scaling path. The down-scaling path extracted information from the aortic annular plane, and the up-scaling path translated this information into a segmented aortic annulus. The final sigmoid activation function transformed the output of the model into probabilities (Figure 5.10).

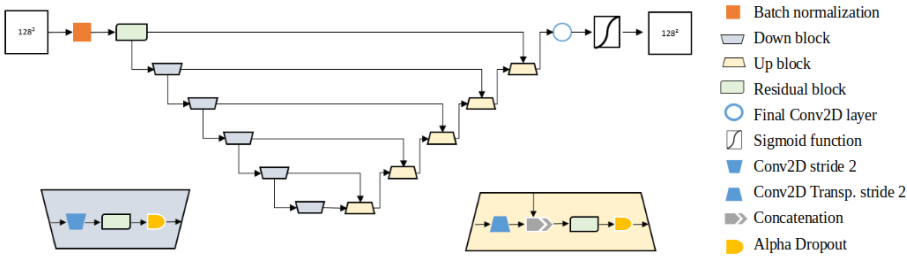



*Figure 5.10: Schematic overview of the architecture used to segment the aortic annulus. The down- and up scaling path of the architecture are defined by the down- and up blocks.*

The down-scaling path consisted of down blocks which decreased the resolution of its input with a convolutional layer (with stride=2). The up-scaling path used up blocks to increase the resolution of its input with a transposed convolutional layer. The down- and up blocks contain a residual block followed by an alpha dropout layer (with p=0.5) [183] which maintained the self-normalising property within the model after performing a dropout.

The residual block introduced an identity shortcut connection, which enables deeper networks without degrading network performance due to poor convergence of learning [52]. After a sequence of convolutional layers with kernel size 3, batch

normalisation [49] and ReLU activation function [43], the output is summed with the output of the first convolutional layer followed by a final ReLU activation function (Figure 5.11).

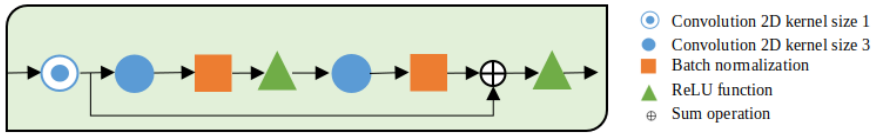



*Figure 5.11: The residual block of the aortic annulus segmentation architecture.*

**Training** Two models were trained using the training dataset and validated with the validation dataset. One model was trained with the 1 mm resolution image data, and the other with the 0.5 mm resolution image data. The validation dataset consisted of the same 118 patients that were used for the inter-observer variability study and the training dataset consisted of the remaining 355 patients. The 36 bicuspid patients were distributed equally over the training and validation dataset. Training details can be found in Table 5.6.

| | | | | | |
|---|---|---|---|---|---|
| **Optimiser** | Adam | **Learning rate** | 1e-2 | **Weight init.** | ROM |
| **Cost** | Dice [184] | **Learning rate decay** | $\frac{1e-2}{\sqrt{epoch}}$ | **Bias init.** | 0.1 |
| **Max. epochs** | 10000 | **Regulariser** | L2 ($\lambda$ = 1e-3) | **K-folding** | k = 5 |

*Table 5.6: Training details for the aortic annular plane segmentation. All hyper-parameters were obtained by performing k-fold cross-validation on the training set (with $k = 5$) and a fixed random seed.*

The following data augmentation strategies were used: random rotation, random value dropout and random calcification placement, each with probability 0.5 of occurring during sample generation. Random rotation rotates the input images with a random angle (sampled between -45 and 45 degrees) to increase the robustness of the model against rotation. Random value dropout fills randomly positioned squares in the aortic annular planes with zeros, which is similar to the working of dropout [46] but was added to control the Hounsfield values of the input before the batch normalisation of the model. Most of the aortic annular planes contained an amount of calcification around the aortic annulus which interfered with the accuracy. Therefore the addition of random calcification in the aortic annular plane was used during training, which was done to increase the robustness of the model for calcified areas. This addition was achieved by filling randomly positioned squares in the aortic annular planes with Hounsfield values that represent calcification. The calcification values were sampled from a normal distribution ($\mu = 1100$ and $\sigma = 50$, which were obtained from the training cohort). Table 5.7 contains the summary of the data-augmentation parameters.

| | |
|---|---|
| **Rotation** | Rotates the input images with a random angle: sampled between -45 and 45 degrees. |
| **Calcification placement** | Add random areas of calcification in the input images: samples from normal distribution $\mu$=1100 and $\sigma$ = 50. |

*Table 5.7: Data-augmentation details for training the aortic annulus segmentation model.*

**Detection** After training one model for each resolution, a detection strategy was used to combine the output of both models and use the segmented aortic annulus to derive patient-specific anatomical information: the area and perimeter of the aortic annulus.

The detection of the area and perimeter of the aortic annulus of a single patient was performed in two steps: a deep learning step and a post-processing step. During the deep learning step, the aortic annular planes were analysed by both models, and the output was combined and normalised to a probability output that identified the region of interest. During the post-processing step, the contour of the region of interest was located with Canny edge detection [185] from the probability output. The area and perimeter were derived from this contour and serve as the final predicted output of the detection phase (Figure 5.12).

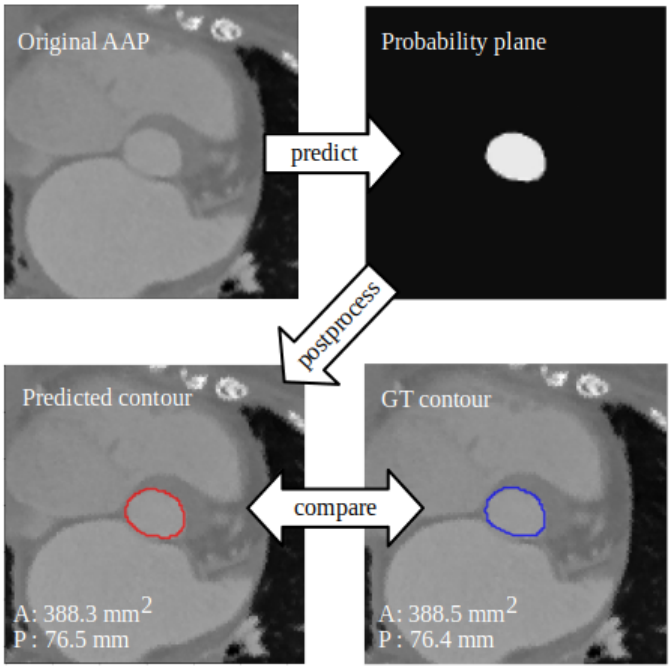



*Figure 5.12: General overview of the method: the model predicts the probability plane from the original aortic annular plane. The contours are detected, and the predicted area and perimeter are compared with the ground truth.*

After the detection phase, the patient-specific anatomical information of the aortic annulus was available. As a final step, this anatomical information was used to assess correct prosthesis size. The perimeter was used to select the proper Medtronic Evolut TAVR prosthesis size, and the area was used to identify the Edwards Sapien 3 prosthesis size similar to the manufacturer's sizing charts used in clinical practice.

#### 5.2.2.4 Statistical Analysis

The Shapiro-Wilk test was performed to test for normal distribution, and none of the predicted distributions were normally distributed. Pearson correlation coefficient was computed to evaluate the correlation between model and both observers (with excellent correlation $R^2 > 0.9$). The agreement between the manual and predicted indices were evaluated using the non-parametric signed Wilcoxon test (with a significant p-value $< 0.05$). Bland-Altman analysis for area and perimeter between model and observer 1 and between both observers was performed.

### 5.2.3 Results

The detection phase was validated using the 118 patients who were also used in the inter-observer variability study. By using the same patients for validation and observer-variability assessment, it was possible to compare the method with both observers.

The detection phase consisted of a deep learning phase and a post-processing phase. The deep learning phase was validated by comparing the predicted segmentation (model) with the segmentation of both observers using the Dice score. The average Dice score between the model and the first observer was 96% whereas the average Dice score between the model and the second observer was 89%. The average Dice score between both observers was also 89%. The higher Dice score between model and observer one is expected because the model was trained with the data from this observer.

The post-processing phase derived the area and perimeter from the predicted segmentation and was validated by comparing the predicted area and perimeter with the area and perimeter of both observers. When comparing the predicted anatomical measurements of the model with the data of both observers, a significant difference between model and both observers can be observed. However, the mean paired difference between model and both observers was around zero, which indicates adequate accuracy of the predicted anatomical measurements (Table 5.8).

| | **Model vs Observer 1** | | **Model vs Observer 2** | | **Observer 1 vs Observer 2** | |
|---|---|---|---|---|---|---|
| | Paired diff. | p-value | Paired diff. | p-value | Paired diff. | p-value |
| **Area ($mm^2$)** | $3.3 \pm 16.8$ | 0.008 | $2.0 \pm 22.4$ | 0.046 | $1.3 \pm 21.1$ | 0.752 |
| **Perimeter (mm)** | $0.6 \pm 1.7$ | 0.0001 | $0.5 \pm 2.6$ | 0.0016 | $0.2 \pm 2.5$ | 0.513 |

*Table 5.8: Comparison of the aortic annulus perimeter and area between model and both observers. Paired difference (Paired diff.) reported as mean ± standard deviation.*

Furthermore, excellent correlation values were obtained between model and observer one for the area (0.98) and perimeter (0.97). The correlation values between observer one and two for the area (0.96) and perimeter (0.94) indicate that the manual method is accurate (Figure 5.13). These results provide additional evidence that the predicted anatomical measurements could be used in the same manner as the output of observer one or two. In Figure 5.14, the Bland-Altman plots of the predicted and measured (observer one) area and perimeter are depicted.

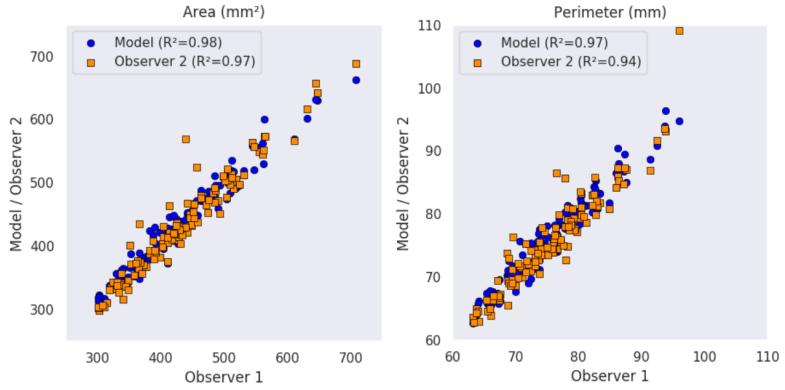



*Figure 5.13: Scatter plots comparing the inter-observer correlation for area (left) and perimeter (right).*

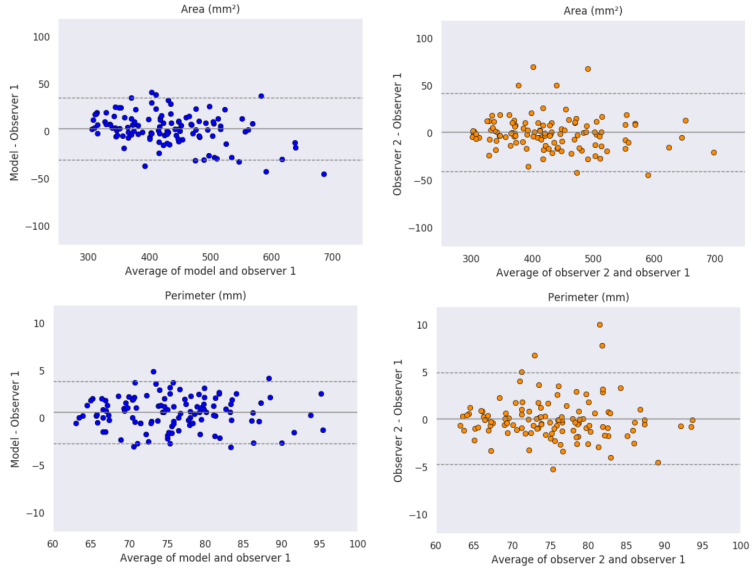



*Figure 5.14: Bland-Altman plots for area and perimeter measurements between observer 1 and the model*

The validation of the segmentation abilities and the area and perimeter assessment were required to validate the method's ability to predict the correct prosthesis size (compared to both observers). The predicted area and perimeter were used to retrieve the Edwards Sapien 3 and Medtronic Evolut TAVR prosthesis sizes. The automatically selected valve sizes were compared with valve sizes resulting from the annular measurements of both observers. The ratio of agreement for Edwards Sapien 3 between model and both observers is equal: 0.86 between model and observer 1 and 0.88 between both observers. The ratio of agreement for the Medtronic Evolut TAVR prosthesis sizes between model and both observers is similar: 0.89 between model and observer 1 and 0.86 between both observers (Figure 5.15).

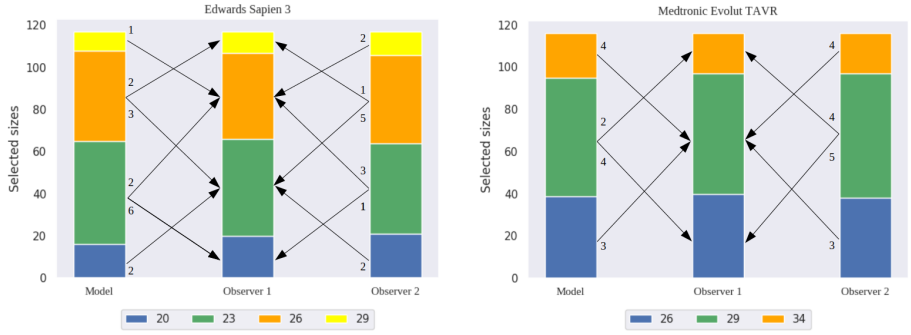



*Figure 5.15: The agreement between prosthesis sizes from the Edwards Sapien 3 (left panel) and Medtronic Evolut TAVR sizing chart (right panel). The plots represent how many sizes were measured of each available device size for model, observer 1 and observer 2. The arrows between the plots indicate disagreement with observer 1 (under- or over-estimation). The weights indicate the number of patients that were missized compared to observer 1.*

After the validation of the predicted prosthesis sizes, it remains to report the processing time of the manual and automated method. Both observers, reported 5 to 10 minutes of analysis time per patient to segment the aortic annulus. The time to derive the prosthesis size from the anatomical measurements is ignored. The automatic processing time from aortic annular plane to segmentation, anatomical measurement and prosthesis size is below 1 second.

### 5.2.4 Discussion

In the literature, similar studies have been conducted. Queirós *et al.* proposed a method for detecting the correct TAVR prosthesis size from the aortic valve annulus area [186]. This method was build on top of aortic segmentation with statistical shape models from previous studies of the same author [187, 188]. In order to detect the correct TAVR prosthesis size, a fully automatic approach and a semi-automatic approach are proposed. Their fully-automatic approach detected 92% of the prosthesis sizes and their semi-automatic approach 100%. This single-center study included 104 patients with a severe degree of calcification, mitral valve prosthesis and pacemakers.

The authors introduced an overlapping area of 35 $mm^2$ and 40 $mm^2$ between the three available prosthesis sizes of the Edwards Sapien 3 and XT. Unfortunately, this overlapping area makes it difficult to assess the real predictive power of the method and to compare with our results. Also, the final processing time was not reported.

Our presented method is based on a different technique and goes, in our opinion, a step further than the work described in Queirós *et al.* Our study includes both aortic valve annulus perimeter and area; therefore, the prosthesis size selection can be expanded to perimeter as well as area dependent devices. Next, multi-center image data was used for training and validating the model, which may indicate robustness to unknown centers. No overlapping region was used in order to follow the manufacturers' guidelines and leave the final interpretation of the output of the method to the physician. Finally, the processing time is around one second per patient, which makes the method efficient.

The method can detect the area and perimeter from the aortic annular plane within seconds, which may have an impact on reducing operator analysis time and errors in an exponentially growing market. In addition to a time reduction of analysis and, thus, procedure planning, the physician saves time as he/she is liberated from this analysis. Also, the analysis concerns an independent automated process that will enhance the output quality. Reduced overall TAVR costs may be obtained by embedding the method in software that allows manual corrections (e.g. to correct outliers). This embedding could also yield a continuous learning platform where the data of a new patient, validated by an expert, can be added to the training dataset, thus improving future detections.

Although the presented method has proven to be reliable, there are a few limitations related to the current approach. In a few cases, relatively significant differences remain between the predicted area from our model and that from an individual human observer. Compared to observer one, the most substantial overestimation of our model amounts to 10% and the most substantial underestimation to 9%. However, in those cases, the second observer tended to agree with the predicted value (1% difference between observer two and the model). This agreement may indicate that the model has generalised beyond the ground truth; in other words, it has learned to look beyond the few inaccuracies of its teacher. The maximum difference between the predicted perimeter and observer one was the same patient as the areas maximum difference (with a 7% overestimation). The minimum difference between predicated area and observer one was a 5% underestimation (a 3 mm difference).

It should be noted that the proposed method is not a TAVR planning tool, nor does it intend to replace the interventional cardiologist. There are other measurements required for the planning of a TAVR which are not included in this study. These measurements include the distance from the aortic annular plane to the ostium of the coronary arteries, the area of Sinus of Valsalva, sino-tubular junction and others and will be addressed in future work. It would also be interesting to measure the impact of this method prospectively.

In the literature, there are no methods that used deep learning to segment a medical image, derive measurements and use those measurements to predict prostheses

sizes. However, as presented in the literature overview, several studies analysed medical images and yielded some quantification [95, 138, 150, 177, 189]. The quantification typically concerned the LV where methods derived the ESV, EDV, EF, LV mass and stroke volume. Some studies predicted the values directly (regression), and other studies derived these values from the segmentation. In this study, we chose the latter approach because we did not know beforehand if it would be feasible to segment the aortic annulus from MDCT images.

Many studies used segmentation with deep learning at the time of this study. However, these studies analysed well-defined areas, e.g. the endocardium of the LV can be delineated visually. In this study, the aortic annulus was segmented. The annular ring also delineates this structure, but calcified areas make the segmentation not straightforward. The operator can include or exclude these calcified areas depending on the size. By segmenting the aortic annulus first, we could establish the confidence that further quantification would yield accurate results. Secondly, we expected that physicians would not appreciate black-box methods. If a model was trained to analyse the MDCT image and yield the area and perimeter as output, it would be difficult to convince a physician how the model came up with these values. Moreover, it would also be challenging to inspect the cause of outliers.

**Conclusion** A method was presented that automatically selects the TAVR device size. The method is efficient, accurate and reproducible. The accuracy was confirmed by comparing the obtained results with the measurements of two independent observers. This accuracy confirmed the reliability of this strategy. Embedding this tool in the pre-operative planning routine may have an impact on the TAVR program by reducing time and cost while ensuring accuracy.

## 5.3 Towards safe and efficient pre-operative planning of TMVR

Transcatheter mitral valve interventions are emerging as a viable alternative for patients at high surgical risk. Two key aspects are crucial during pre-operative planning: left ventricular outflow tract (LVOT) assessment and anatomical analysis. Given that the manual anatomical analysis is time-consuming, an automated approach may introduce efficiency during pre-operative planning. In this study, an automatic method is presented to detect the mitral valve annulus, and the possible implementation of this method in clinical practice is discussed.

### 5.3.1 Introduction

Since transcatheter mitral valve replacement (TMVR) is a closed-chest procedure, pre-operative imaging is of paramount importance for identifying patients eligible for TMVR and optimal pre-operative planning. A life-threatening and relatively frequent complication after TMVR is an obstruction the LVOT, so it is crucial to assess the risk on this complication pre-operatively. A simple overlay of an undeformed device

on the computed tomography (CT) images can help to understand this risk [190, 191], or a more detailed analysis involving computational modelling can be performed [192]. When a patient is considered eligible for TMVR, a detailed anatomical analysis needs to be performed, for example, to determine device size. This analysis is not a trivial task due to the complexity of the mitral valve apparatus. For example, the mitral valve annulus is a non-planar structure of which the dimensions can significantly vary throughout the cardiac cycle. Identifying this structure is, therefore, a time-consuming process.

In this work, we show how deep learning can be used to detect the mitral valve annulus automatically with excellent efficiency. Next, we evaluate the accuracy of the derived mitral annulus dimensions and finally, a possible implementation of this technology in clinical practice is presented and discussed.

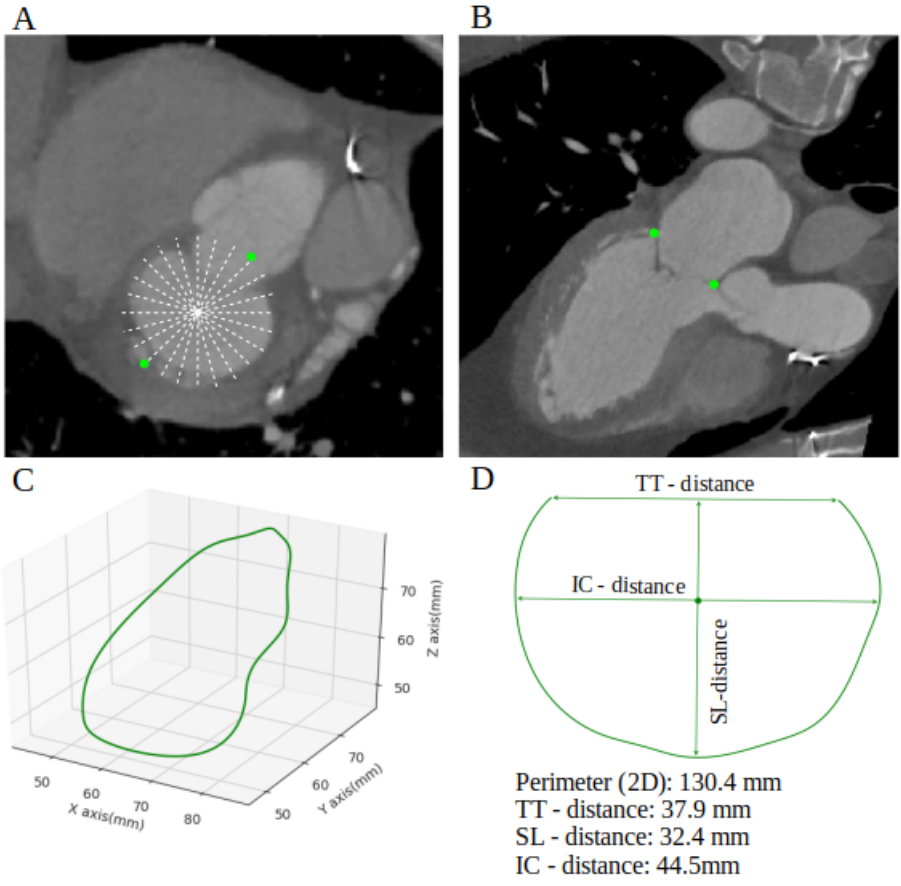



*Figure 5.16: Overview of the procedure to manually detect the mitral valve annulus. First, the 24 points of the mitral valve annulus were detected (A, B). Next, the 24 points were interpolated into a closed 3D curve (C). Finally, a 2D curve was generated from which the clinical measurements were derived (D).*

## 5.3.2 Materials and methods

### 5.3.2.1 MDCT images

This retrospective study used the data of 71 patients collected from multiple centra. The average age of this cohort was $74.2 \pm 13.1$ years, and 43.9% of the patients were male and 56.1% female. The patient data consisted of a contrast-enhanced volumetric MDCT image that was recorded at the end-systolic phase of the cardiac cycle. The average slice thickness of the MDCT images was 0.7 mm. The entire dataset was used in both manual and automatic detection.

### 5.3.2.2 Manual detection

The mitral valve annulus was detected in the volumetric MDCT images for all patients by a trained operator. This person manually detected 24 mitral annular points and the left- and right trigone for the entire cohort with the Mimics Innovation Suite 18 (Materialise, Leuven, Belgium). A second expert inspected all annotations during a quality control check.

First, the center of the planar part of the mitral annulus was identified, and the coronal axis was aligned with the valve in the axial and sagittal views. Then, in both the axial and sagittal view, the edges of the mitral valve were identified, and four points were marked. Next, the sagittal axis in the coronal view was rotated 15 degrees clockwise, and the process of picking the points in the axial and sagittal views was repeated. This process was repeated four more times, which resulted in a total of 24 points (4 points in angles 0, 15, 30, 45, 60, 75). The left- and right trigone points were detected using similar reslicing planes (Figure 5.16 A and B).

The mitral valve annulus and trigone points were considered the ground truth in this study. The manual segmentation of the mitral valve annulus was considered a robust yet time-consuming procedure.

### 5.3.2.3 Automatic detection

In order to automate the manual detection of the mitral valve annulus from MDCT images, pre-processing of the MDCT images and the ground truth was required in order to prepare the data for the deep learning models.

**Pre-processing** The volumetric MDCT images were resampled to an isotropic 1 mm resolution to ensure a homogeneous dataset. The original Hounsfield units were preserved during the resampling process, which was performed with cubic spline interpolation. The images were centered around the center of the mitral valve annulus points and clipped to $128^3$ voxel cubes to fit the input of the model.

The ground truth annotations needed to be transformed into volumetric binary masks in order to teach the deep learning models how to segment the mitral valve annulus. These masks had the same dimensions as the MDCT images and contained the information of the area around the mitral valve annulus. The transformation for a single patient was performed in three steps. First, the 24 ground truth coordinates

were transformed into a closed 3D curve which was smoothed by applying b-spline interpolation (Figure 5.16 C). Next, the smoothed curve was used as centerline to construct a 3D cylinder around it. Finally, the 3D cylinder was positioned in the volumetric binary mask to annotate the area around the mitral valve annulus. In order to increase the level of granularity, different binary masks with different radii (2, 3 and 5mm) for the 3D cylinders were exported per patient. The binary masks were used to train three deep learning models (one per radius) that could segment the mitral valve annulus with a different degree of precision. The time to pre-process the data of a single patient was below one minute.

**Architecture** In order to focus on the clinical part of this study, an already established deep learning architecture was chosen. The DenseVNet architecture [26] is an image-to-image deep learning segmentation architecture which is composed of a downsampling path followed by an upsampling path yielding its v-shape. The downsampling path reduces the resolution of the input image to low-resolution representations by using strided convolutions, and the upsampling path increases the output of the downsampling path to the original dimensions by using bilinear upsampling. An input image x flows from the downsampling path to the upsampling path which is followed by the non-linear softmax activation function which generates the segmentation image $\hat{y}$. During training and detection, an input image x is analysed by the network, and the segmentation image $\hat{y}$ is produced as output.

#### 5.3.2.4 Training and validation

Three models were trained and cross-validated using k-folding with k=4. The entire dataset was split into four subsets, and for each fold, a different subset was excluded from the training dataset and used as the validation dataset. With k-folding, the limitation of a small dataset was evaded, and the entire cohort was used to validate the method.

One model was trained for each of the three radii (2, 3 and 5 mm) defining the mitral valve annulus in the MDCT images. In order to improve the convergence speed of the training process, the MDCT images were normalised using feature standardisation [49]. The training details are summarised in Table 5.9. The following data augmentation strategies were used: random rotation and random scaling. Random rotation (in degrees) and random scaling (in percentage) was performed by sampling values from a uniform distribution ranging between -3 and 3.

| **Optimiser** | Adam | **Learning rate** | 1e-4 | **Cost** | Dice [26] |
|---|---|---|---|---|---|
| **Weight init.** | ROM with gain = $\sqrt{2}$ | **Bias init.** | 0.1 | **Regulariser** | L2 ($\lambda$ = 1e-4) |

*Table 5.9: Training details of the mitral valve annulus segmentation training. All hyper-parameters were obtained by performing k-fold cross-validation on the training set (with k = 4) and a fixed random seed.*

#### 5.3.2.5 Detection

The detection of the mitral valve annulus of a single patient was performed in two steps: a deep learning step and a post-processing step.

During the deep learning step, a medical image was segmented as follows. A volumetric MDCT image was analysed by the three trained models (or ensemble [116]). The output of these models had the same dimensions as the MDCT image and contained information where the deep learning models detected the area around the mitral valve annulus. These three output volumes were combined and normalised into a volumetric probability output (Figure 5.17 A).

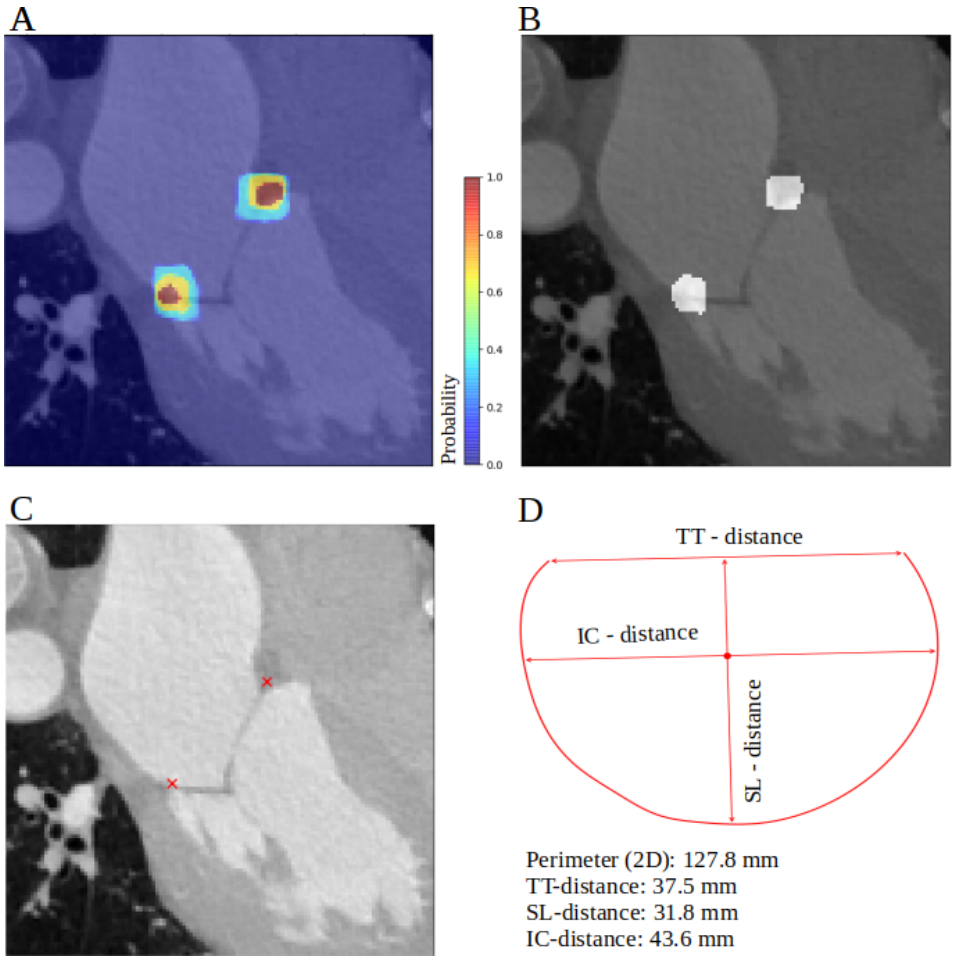



*Figure 5.17: Overview of the automatic mitral annulus detection strategy. The output of the deep learning models was normalised into one output mask (A). Next, the decision boundary was applied to create a binary mask (B), after which the closed 3D curve was extracted (C). Finally, the clinical measurements were derived from a 2D curve (D).*

During the post-processing step, a closed 3D curve was extracted from the probability volume as follows. First, the probability volume was transformed into a binary mask by using a decision boundary of 0.5 (Figure 5.17 B). Next, in order to obtain a 3D centerline of the detected area around the mitral valve annulus, the 3D medial surface and axis thinning algorithm [193] was used. This algorithm yields a 3D skeleton from a binary image. In case of a perfect detection, this 3D skeleton would form a closed 3D curve. However, due to occasional outliers in the segmentation, this skeleton could contain additional non-circular paths attached to it. This issue was addressed in the final step, where the largest circular path in this skeleton was obtained as follows. The 3D skeleton was transformed into a bidirectional graph by using the 3D coordinates of the skeleton as nodes and all neighbouring nodes within a radius of 3 mm were added as edges. The biggest cycle in this graph was selected using the algorithm described in [194]. This cycle or closed 3D curve was considered the predicted mitral valve annulus in this study (Figure 5.17 C).

As a final step, clinical measurements were determined from the 2D posterior points of the predicted mitral valve annulus. The posterior points were selected and projected onto a two dimensional (2D) plane using principal component analysis (PCA). The predicted 2D posterior points were used in order to obtain the following clinical measurements: trigone-to-trigone (TT)-, septal-to-lateral (SL)-, commissure-to-commissure (IC)-distance and the centroid and perimeter of the 2D posterior points [191] (Figure 5.17 D).

#### 5.3.2.6 Metrics

In order to compare predicted and ground truth closed 3D curves with each other, the Hausdorff distance was used. This metric measures the greatest of all distances from one point in a set to the closest point in another set.

### 5.3.3 Results

The detection phase consisted of a deep learning phase and a post-processing phase. The deep learning phase was validated by measuring the Dice score between the prediction and ground truth. This validation yielded an average Dice score of 74% between the predicted segmentation (model) and ground truth binary masks. After the deep learning phase, the post-processing phase was responsible for extracting the predicted mitral valve annulus. The paired difference between predicted and ground truth closed 3D curves was $4.57 \pm 1.64$ mm (Hausdorff distance). This difference is acceptable since Hausdorff represents the maximal distance between the ground truth and the predicted closed 3D curve and thus reveals that the predictions are close to the ground truth. A visual comparison between predicted and ground truth mitral valve annulus is presented in Figure 5.18.

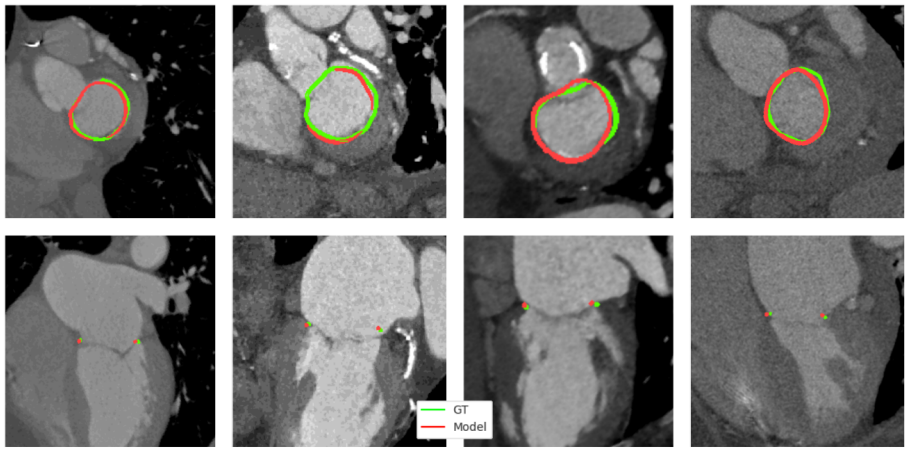



*Figure 5.18: A visual comparison between ground truth (GT) and prediction (Model) of 4 patients. The mitral valve annular plane (top) and the sagittal view (bottom) are depicted. The mitral valve annulus of the prediction is presented in red, and the ground truth is presented in green.*

Following the validation of the segmentation and the closed 3D curve, clinically relevant anatomical measurements were derived from the predicted mitral valve annulus, using the posterior part of the mitral valve annulus. The predicted 2D and 3D perimeter were compared to the ground truth and displayed excellent correlation despite a slight overestimation of these predictions. Similarly, the predicted TT-, SL- and IC-distance were compared to the ground truth, showing a strong correlation between the predicted- and ground truth distances (Table 5.10).

| **Observer vs. Model** | | | |
|---|---|---|---|
| **Measurement [mm]** | **Paired diff. [mm]** | **p-value** | **$R^2$** |
| 3D Perimeter | $-2.4 \pm 6.3$ | $<0.001$ | 0.92 |
| 2D perimeter | $-1.2 \pm 4.6$ | 0.02 | 0.93 |
| TT-distance | $0.12 \pm 2.56$ | 0.5 | 0.86 |
| SL-distance | $-0.32 \pm 2.19$ | 0.2 | 0.86 |
| IC-distance | $-0.06 \pm 1.88$ | 0.5 | 0.90 |

*Table 5.10: Comparison between the clinical measurements of the observer and the predicted measurements. Paired difference (Paired diff.) reported as mean ± standard deviation.*

Bland-Altman and correlation plots of the measurements of the 2D perimeter are depicted in Figure 5.19.

Besides the accuracy of the anatomical measurements, it is also essential to report the processing time of the manual and automated method. The trained operator reported approximately 25 minutes of processing time per segmentation of a single phase of the cardiac cycle. The automatic processing time from single pre-processed MDCT image to the predicted segmentation and mitral valve annulus with 2D clinical

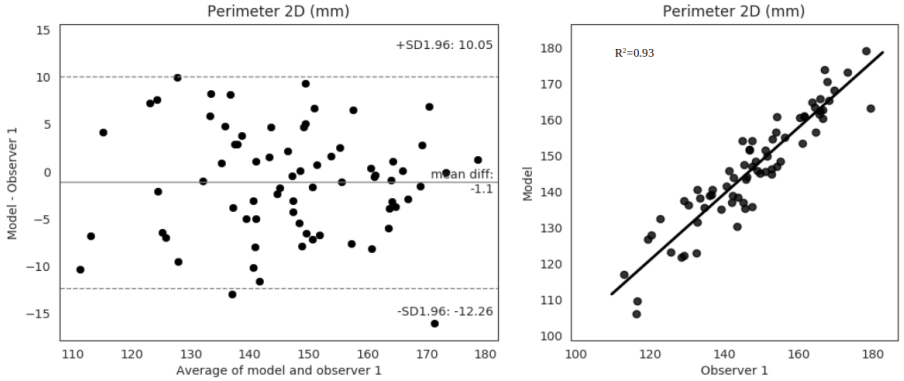



*Figure 5.19: Bland-Altman and correlation plot comparing the 2D perimeter of the predicted mitral valve annulus (Model) and ground truth (Observer 1).*

measurements was below 1 second which is a speed-up of a factor of 1500 compared to the manual analysis time.

### 5.3.4 Discussion

Treating mitral regurgitation patients through transcatheter mitral valve replacement has enormous potential. However, early experience clearly showed that detailed screening and planning are required to minimise the risk of life-threatening complications [19]. Two crucial aspects of this are the assessment of the risk of LVOT obstruction and detailed quantification of the mitral valve anatomy. In this study, a method to automate the latter was proposed.

One way of introducing the proposed automatic anatomical quantification into clinical practice is through a web-based service model. Pre-operative MDCT images are uploaded to a web platform, triggering the automatic anatomical analysis of the images as described in this work. As with any prediction model, the automated annulus detection may lead to inaccurate results, mainly when the images of a new patient differ significantly from the data on which the deep learning model has been trained. Therefore, it is vital that the workflow contains human quality control (QC) and allows manual corrections. During a final step, a report is generated. Since the anatomical detection and derived measurements are verified by an expert, the MDCT image and anatomical data can be used to increase the dataset and thus to yield a continuous learning platform (Figure 5.20).

As mentioned above, the assessment of the risk of left ventricular outflow tract obstruction is another crucial aspect when considering transcatheter mitral valve replacement. A simple overlay of an undeformed device on the CT images [190, 191] or computational modelling [192] have been proposed to evaluate the risk of left ventricular outflow tract obstruction. These analyses could be perfectly integrated into a web-based service.

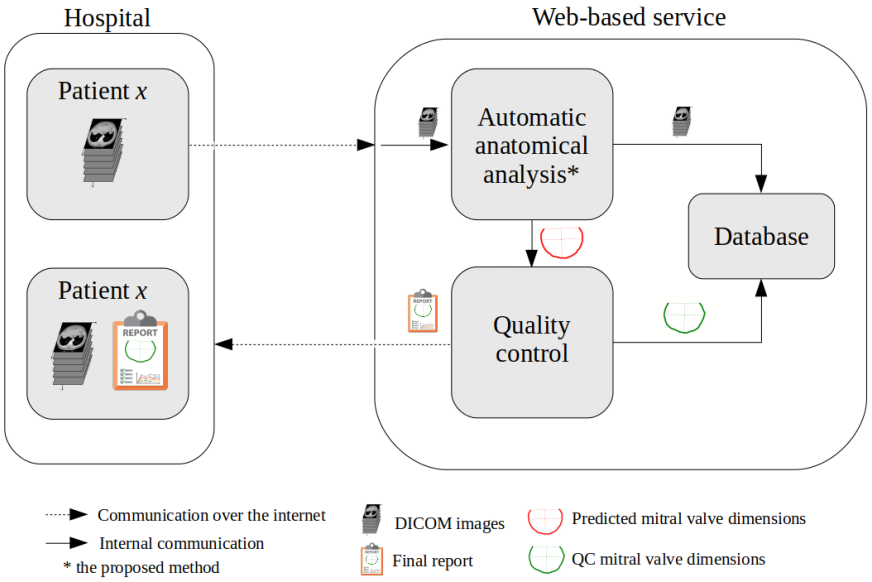



*Figure 5.20: Schematic overview of the implementation of the proposed method in a clinical workflow for planning transcatheter mitral valve interventions*

In the literature, few studies have been conducted to detect (parts of) the mitral valve apparatus. Voigt *et al.* presented an algorithm to detect and track the mitral valve annulus over long image sequences. This study was validated on twelve 4D transesophageal echocardiography (TEE) images [195]. In the work of Stralen *et al.*, a mitral valve plane detection algorithm from 3D echocardiograph was presented and validated on 25 patients [196]. In the work of Schneider *et al.*, 3D graph cutting was applied to segment the mitral valve leaflets from the 4D ultrasound [20]. Pouch *et al.* proposed a semi-automatic mitral valve detection method based on shape-fitting from 3D TEE images [21]. Pedrosa *et al.* proposed a fully automatic mitral valve morphology detection using a b-spline explicit active surfaces framework from 3D TEE images of the end-systolic and end-diastolic phase. A cohort of 15 patients was analysed and the validation of the automatic morphology detection yielded a Hausdorff distance of $6.70 \pm 1.97$ mm at end-systolic phase and $5.79 \pm 1.25$ mm at the end-diastolic phase [197].

The studies mentioned above propose methods to detect (parts of) the mitral valve apparatus from 3D or 4D TEE images. The validation of these studies was limited to a few patients but yielded accurate results. To the best of our knowledge, our study is the first that detects the mitral valve annulus and the clinical measurements thereof from MDCT images, including a validation on 71 patients which is higher than the related state-of-the-art.

The accuracy of this method is confirmed by comparing the obtained measurements to the work of Mak *et al.* [198]. In this study, the authors compared two image modalities by analysing 41 patients and observed a good correlation between

measurements made with 3D TEE and 3D CT images of the mitral valve annulus. They report a mean 2D mitral annulus circumference from CT of $123.9 \pm 15.5$ mm which is similar to the mean circumference of our findings: $147.97 \pm 14.96$ mm (ground truth) and $146.40 \pm 15.21$ mm (model). They also reported mean SL- and TT-distances of $32.5 \pm 4.4$ mm and $32.6 \pm 3.6$ mm which is comparable to our findings: $28.5 \pm 4.2$ mm vs $28.2 \pm 3.9$ mm (model) and $32.0 \pm 4.8$ mm vs $32.2 \pm 4.7$ mm (model).

Although we showed the efficiency and accuracy of this method, there are a few limitations related to the current approach. First, the deep learning models expect a volume that is centered around the mitral valve annulus. Providing this location may introduce a bias given that the original DICOM may range from full-body scans to ROI images around the left ventricle and atrium. This bias can be resolved by introducing a pre-processing landmark- or ROI-detection model before the proposed method. Secondly, the size of the dataset used in this work is small. This size limitation was handled by adding data-augmentation during the training and using k-folding to cross-validate the method appropriately. Adding more patients to the cohort would benefit the robustness of the proposed method, which could be obtained by enabling the continuous learning platform, as previously explained. Finally, transcatheter mitral valve intervention pre-operative planning is not limited to a single phase of the cardiac cycle. Currently, the models are trained with MDCT images of the end-systolic phase, and the extension of the detection for the full cardiac cycle will be tackled as a next research step.

In this work, we demonstrated a method that can detect the mitral valve annulus and derive clinically relevant measurements automatically from MDCT images. In the presented literature study, there are no studies that use deep learning to segment a medical image to derive the mitral valve annulus and secondary measurements thereof. There are a few studies that use deep learning and other methods to detect the contours of the LV [130, 199]. The scope of our work targets a less-defined anatomical structure, i.e. the LV is a well-delineated on CT images whereas the mitral valve annulus is not.

**Conclusion** A method was presented that detects the mitral valve annulus and the clinical measurements thereof in an efficient, accurate and automatic way. A possible implementation of this method in clinical practice was proposed and discussed that moves us closer towards safe and efficient pre-operative planning of transcatheter mitral valve interventions.

## 5.4 Conclusion

In the first part of this chapter, a literature study of cardiac segmentation with deep learning between 2012 and 2019 was presented. In this overview, the evolution of deep learning over the past years is made clear: starting from "simple" CNN models towards a standardised method for segmentation.

In the second part of this chapter, two studies which used segmentation with deep learning were presented. The first study described a method to automatically

determine the TAVR device size from the aortic annular plane. The second study segmented the mitral valve annulus from volumetric MDCT images and derived clinical measurements from the annulus. Both presented methods were efficient and accurate, which could benefit the transcatheter aortic- and mitral valve solutions.

# 6
# Landmark detection

*Needle in a haystack is easy,*
*just bring a magnet*

Keith R.A. DeCandido

Landmark detection is the recognition of a specific feature in a defined search space. This definition is similar to the definition of segmentation but the difference between these two is that landmark detection deals with very small solutions when compared to the available search space. This chapter starts with a literature overview of landmark detection in medical images with deep learning. Subsequently, two different approaches of landmark detection will be presented and applied to detect three-dimensional (3D) landmarks, which are essential during the transcatheter aortic valve replacement (TAVR) pre-operative planning.

## 6.1 Literature overview of landmark detection with deep learning

In this section, a literature overview of landmark detection is presented. This literature study is limited to landmark detection in medical images with deep learning. As the number of papers is limited, also non-cardiovascular work was considered.

### 6.1.1 2015-2019

This literature overview of landmark detection with deep learning in medical images starts with a 'simple' deep neural network model in 2015 [200]. Note that at this point, methods and architectures such as AlexNet, U-Net and batch normalisation

were already established. Other studies used 3D CNN models [201–203] or even deep reinforcement learning methods such as deep q-learning (DQL) [204]. Table 6.1 contains more information about the discussed work.

| Authors | Scope | Year | Patients | Modality | Method | Dim. |
|---|---|---|---|---|---|---|
| Zheng *et al.* [200] | DNN approach | 2015 | 455 | CT | FC | 3D |
| Payer *et al.* [205] | Regressing heatmaps | 2016 | 895 | CT/MR | CNN | 2/3D |
| Zhang *et al.* [189] | Displacement vector with coordinate prediction | 2017 | 700 | MR | CNN | 3D |
| Xue *et al.* [206] | LV indices prediction | 2017 | 73 | CT | CNN | 3D |
| Ghesu *et al.* [204] | DQL agent path prediction | 2017 | 532 | CT | RL | 3D |
| Al *et al.* [201] | Colonial walk | 2018 | 71 | CT | Regr | 3D |
| O'Neil *et al.* [202] | Regressing heatmap and landmark atlas | 2018 | 201 | CT | CNN | 3D |
| Noothout *et al.* [203] | Patch-based displacement vector prediction | 2018 | 198 | CT | CNN | 3D |
| Dangi *et al.* [207] | LV short-axis re-alignment | 2018 | 97 | MR | CNN | 2D |

*Table 6.1: Summary of landmark detection with deep learning in 2015-2019*

#### 6.1.1.1 Detailed literature overview

In Zheng *et al.*, a deep neural network (DNN) landmark detection strategy was proposed and applied to detect the carotid artery bifurcation from computed tomography (CT) images. The authors used a shallow neural network to identify 2000 candidate image-patches, which may include these landmarks. A second, three-layered neural network was used to remove the false positives and detect the target landmarks. The authors increased the accuracy of their method by using an image pyramid with three resolutions and using an ensemble of different models. The final accuracy obtained was a mean error of 2.64 mm with a detection speed of 1 second per volume [200].

A novel method was proposed by Payer *et al.*, where a single model was used to detect bone landmarks from magnetic resonance (MR) and CT images in hands. In this method, heatmaps were used to represent the location of the landmarks. The method was applied to a two-dimensional (2D) CT and 3D MR dataset of landmarks annotated by experts. Four types of models were evaluated: a convolutional neural network (CNN) model, a fully convolutional neural network (FCNN) model, U-Net and a custom model that used channel-wise convolutions. The authors concluded that the U-Net model was slightly better than other approaches with an average error of 0.87 mm in 2D and 1.18 mm in 3D [205].

Zhang *et al.* proposed a two-step approach to detect 3D landmarks in MR images of the brain and CT images of the prostate. The first step consisted of a CNN model and a patch-wise training strategy where the model learned a displacement vector map from the input patches to the target landmarks. The cost function was adapted to give preference to nearby landmarks. In the second step, the weights of the first-stage model were used and frozen. The architecture was expanded to predict the coordinates of the landmarks directly. This method depended on a 3D linear registration of the entire dataset before any learning in order to obtain the anatomical correlation between the input images. The average error that was obtained from the CT images was 3.34 mm, and the detection speed was 1 second per volume [189].

Similar to Zhang *et al.*, a 3D patch-based displacement vector mapping strategy was proposed by Noothout *et al.* The CNN model was a six-layered CNN model followed by two fully-connected layers and concluded with two outputs. This model was trained to detect the 3D displacement vector from image patches (Figure 6.1). The model was also trained to classify these patches whether or not they included the target landmarks. The method was applied to detect six TAVR-related landmarks from CT images. The target points included the aortic ostia (2), the aortic commissure points (3) and the bifurcation of the left main coronary artery (1). The overall average error was 2.44 mm [203].

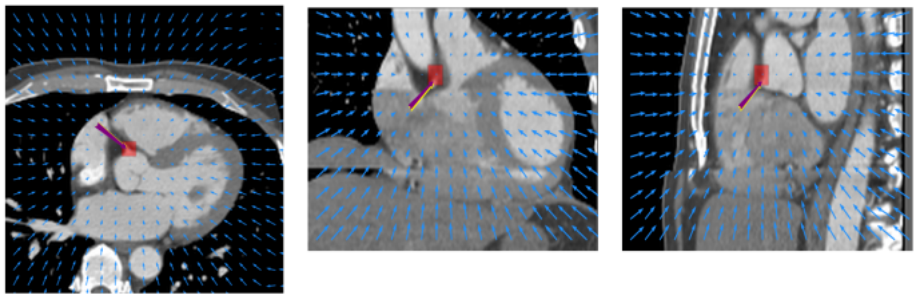


*Figure 6.1: The vector field prediction used to trace TAVR landmarks. Image from [203].*

In Xue *et al.*, a method to predict cardiac indices from MR images was proposed. The method used two CNN models: an image-to-image auto-encoder and a multiple-output CNN. The first model was used for cardiac image representation and the second model used the output of the auto-encoder to predict six indices of the left ventricle (LV) wall thickness, the cavity area of the LV and the area of the myocardium. The average error distance of the linear indices was $1.44 \pm 0.71$ mm [206].

A different approach was proposed by Ghesu *et al.*, where deep q-learning (DQL) was used to detect the target landmarks. An agent was trained to find the most optimal path from anywhere within the 3D CT images to the target landmarks. The method used a scaling technique to overcome the vast search area introduced by full-body scans. The method was applied to many target landmarks, with the most relevant one being the left subclavian artery bifurcation where the average error was 3.09 mm with a reported median run-time of 33 ms [204].

Al *et al.* proposed a regression tree-based machine-learning algorithm and applied it to detect 3D landmarks in CT images. The eight target landmarks consisted of TAVR-related landmarks: aortic valve hinge points (3), the commissure points (3) and the coronary ostia (2). The reported average distance error was 2.09 mm and was obtained by applying the method to non-TAVR patients. The reported average run-time per patients was 12 ms [201].

Similar to Payer *et al.*, a two-step method to detect landmarks using regression to Gaussian points with a CNN model was proposed by O'Neil *et al.* The six-layer 3D

CNN model (concluded by a linear activation function) was trained with a weighted mean squared error (MSE) cost function. The first step resulted in proposals for the landmarks which were transformed into atlas coordinate maps. The second pass used these in combination with the input image from the first pass. The method was applied to detect 22 landmarks from head CT scans, which yielded an average error of 3.38 mm and 2.93 mm after the first and second pass [202].

In Dangi *et al.*, a method to re-align short-axis MR images during segmentation was proposed. The CNN model was trained to predict the center of the LV in the slices directly. The method was validated on nine patients, and the slice misalignment was reduced from a median of 3 pixels to a median of 2 pixels [207].

### 6.1.2 Conclusion

This concludes the literature overview of landmark detection with deep learning. In Figure 6.2, a timeline with milestones in this literature overview and the research in this chapter are depicted. In the next sections, two different methods that detect landmarks from medical images are presented.

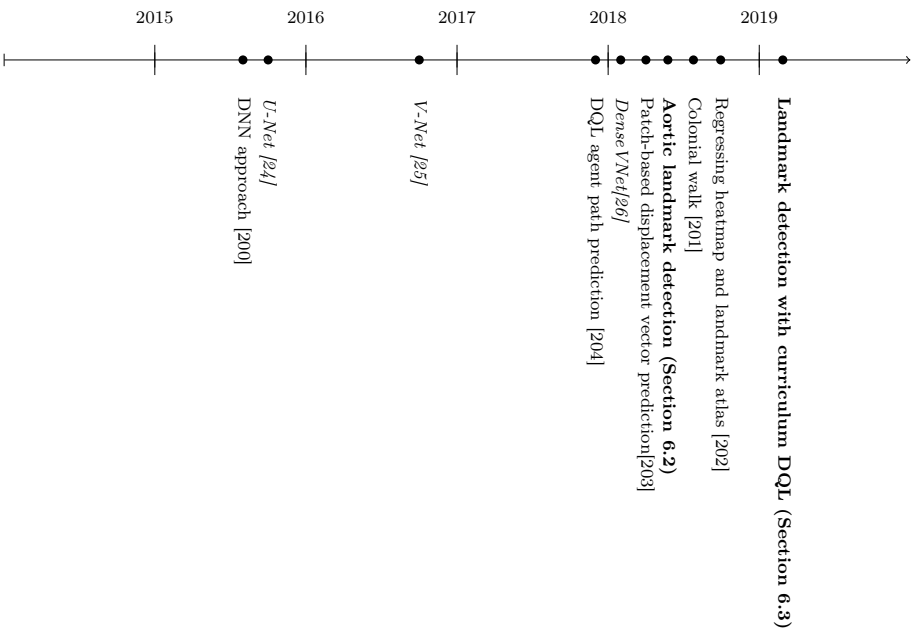



*Figure 6.2: Chronological overview of the most significant milestones of the presented literature, the aortic landmark detection study and landmark detection with curriculum deep Q-learning.*

## 6.2 Aortic annular plane and coronary ostia detection

### 6.2.1 Introduction

As we already saw in the previous chapter (Section 5.2), the selection of the optimal aortic prosthetic valve size is crucial for the short- and long term success of a TAVR procedure. Incorrect sizing may lead to adverse events: over-sizing may cause aortic annulus rupture, coronary obstruction or conduction abnormalities while under-sizing may increase the risk for paravalvular regurgitation or device migration.

The aortic annulus perimeter, area or diameters are measured at the aortic annular plane (AAP), which is defined by the three basal attachment points of the aortic valve leaflets: the left coronary cusp (LCC), non coronary cusp (NCC) and right coronary cusp (RCC). The pre-operative MDCT images are also used to identify patients at risk for other complications. For example, the distances from the AAP to the left coronary ostium (LCO) and the right coronary ostium (RCO) are typically measured to understand the risk of coronary obstruction, a potentially life-threatening complication during which the blood flow to a coronary artery is significantly reduced. Detecting the aforementioned landmarks (LCC, NCC, RCC, LCO, RCO) on MDCT using the manual method [182] is time-consuming and both accuracy and reproducibility are strongly dependent on the operator experience and image quality.

TAVR is rapidly expanding towards intermediate and low-risk patients. Therefore the need for a fast, automatic, reproducible and accurate method becomes even more critical as optimal sizing and adequate risk assessment are of paramount importance for these patient groups. An automatic method that fulfils the above criteria has the potential to speed up the pre-operative planning and to improve device and patient selection, thereby reducing costs and the risk of procedural failure.

In this study, an automated method to extract landmarks for pre-operative TAVR planning from MDCT using deep learning techniques is proposed. The accuracy of the proposed method are assessed using a cohort of 100 patients. The results are compared with an inter-observer variability study on the same 100 patients.

### 6.2.2 Materials and methods

#### 6.2.2.1 MDCT imaging

This retrospective study used the data of 444 patients collected from multiple centra. The average age of this cohort was 80.8 $\pm$ 7.1 years, and 55% of the patients were female and 45% male. The patient data consisted of volumetric MDCT images, which were acquired to support the pre-operative phase of a TAVR procedure. Therefore, all MDCT images were contrast-enhanced and contained a certain degree of aortic stenosis. The average row, column and slice thickness of the MDCT images were 512.1, 511.7 and 0.8 mm. The MDCT images were collected from multiple hospitals, which introduced a variety in recording methods and image qualities. Some images showed motion artefacts due to cardiac motion, whereas others displayed metallic artefacts

due to the presence of medical devices in the patient. Few images contained regions of noise with no cardiac information.

#### 6.2.2.2 Manual detection

A trained operator manually detected the five landmarks from the entire cohort with Materialise Mimics Innovation Suite 18 (Mimics, Leuven, Belgium). The LCC, NCC and RCC were detected following the guidelines [182]. First, the center of the aortic root was detected and used to align the longitudinal axis of the coronal and sagittal plane. Next, the transverse plane was aligned at the level of the valve. The plane was lowered from the aorta towards the ventricle until the basal attachment points of the aortic valve leaflets were visible in this view. Depending on the orientation of the aortic plane, all three basal attachment points need to disappear at the same time; otherwise, re-orientation is required. This re-orientation proved to be a difficult task when patients presented with a high calcium load or when the image quality was not sufficient to identify the basal attachment points. The LCO and RCO were visually detected. In order to measure the coronary heights, the AAP was reconstructed from the three basal attachment points. The distances from the AAP to the LCO and RCO were measured. These five manually detected landmarks were used as ground truth in this study. Figure 6.3 depicts an overview of the landmarks.

A second trained operator re-detected the five landmarks blindly for 100 randomly selected patients using the same method as mentioned earlier. This data was used to assess the inter-observer variability. The same 100 randomly selected patients were used during the validation of the proposed method.

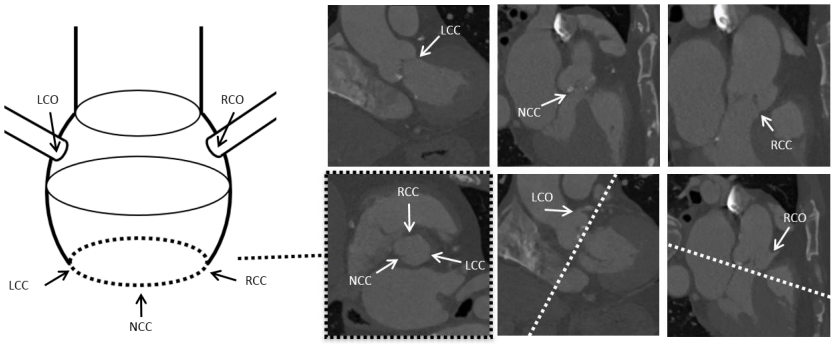



*Figure 6.3: The schematic representation of the anatomy of the aortic root (left). The top three images show the three basal attachment points: LCC, NCC and RCC (from left to right). The bottom three images are (from left to right) the AAP (with the three basal attachment points), the LCO (with AAP, dashed line) and the RCO (with AAP, dashed line).*

#### 6.2.2.3 Automatic detection

The automatic detection of 3D landmarks in MDCT images can be a difficult task due to the image quality and size variation. Methods to detect landmarks from medical images have already been proposed [137, 189, 200–202, 205, 208], and although these methods showed promising results, their coarse accuracy or image modalities did not apply to this problem. Therefore, a heatmap detection method similar to Payer *et al.* was proposed with varying heatmap sizes as an incremental novelty.

This study focused on automating the manual landmark detection which required the pre-processing of the ground truth data and MDCT images.

**Pre-processing** The volumetric MDCT images were resampled in order to obtain a homogeneous dataset. Each MDCT image was resampled to isotropic resolutions of 1.0 mm. The original Hounsfield units (HU) were preserved during the resampling process, which was performed with cubic spline interpolation. Next, a $128^3$-voxel volume centered around the ground truth landmarks was extracted from the resampled volumes. The centering was required because the scanned regions ranged from aortic root-specific to entire body scans (Figure 6.4). In order to improve the convergence speed of the training process, the MDCT images were normalised using feature standardisation [49].

Masks were generated in order to teach the model where to find the target landmarks. A zero-valued $128^3$-cube containing five spheres centered at the location of the manually detected landmark in the associated MDCT image. Each sphere was assigned with its own value: LCC = 1, NCC = 2, RCC = 3, LCO = 4, RCO = 5 (Figure 6.4). In order to increase the accuracy, masks with three sphere radii (3, 5 and 7 mm) were exported.

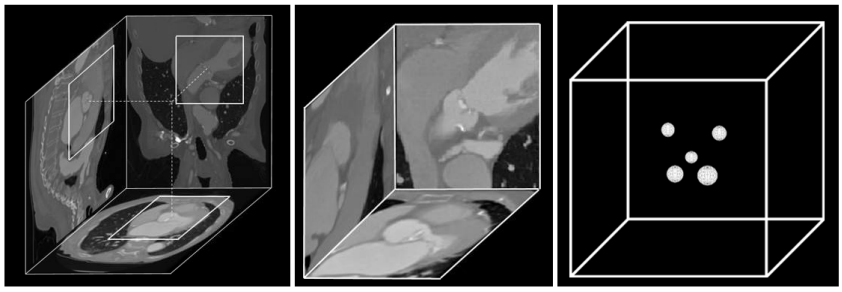


*Figure 6.4: Overview of the MDCT images and the created mask. The extraction of a cube from the full MDCT image (left). An interpolated $128^3$-voxel volume centered around the aortic valve (center). The mask: a zero-valued cube containing five uniquely-valued spheres centered around their associated landmark (right).*

**Architecture** In order to focus on the clinical part of this study, an already established image-to-image DenseVNet architecture [26] was chosen. An input image flows from the downsampling path to the upsampling path followed by the non-linear softmax activation function, which generates the probability segmentation image.

**Training** Three models (N3, N5 and N7) were trained using the training dataset and validated using the validation dataset. One model was trained for each sphere radius (3, 5 and 7 mm). The validation dataset consisted of the same 100 patients that were used for the inter-observer variability study and the training dataset consisted of the remaining 344 patients. Table 6.2 contains the summary of the training details.

| | | | | | |
|---|---|---|---|---|---|
| **Optimiser** | Adam | **Learning rate** | 1e-4 | **Input size** | $64^3$ |
| **Weight init.** | ROM with gain = $\sqrt{2}$ | **Bias init.** | 0.1 | **Nr. of epochs** | 2500 |
| **Cost** | Dice [25] | **Regulariser** | L2 ($\lambda$ = 1e-4) | | |

*Table 6.2: The training details of the aortic landmark detection training. All hyper-parameters were obtained by performing k-fold cross-validation on the training set (with k = 5) and a fixed random seed.*

Random image transformations and randomised window clipping were used as a data-augmentation strategy. The values for random scaling (mm) and rotating (degrees) were sampled from a uniform distribution ranging between -3 and 3. The random transformations and window clipping decreased the dependency upon the ground truth when centering the $128^3$-voxel MDCT volumes around the five landmarks in the pre-processing step.

**Detection** A detection strategy was used to combine the output of the three trained models and use the five predicted landmarks to derive patient-specific anatomical information.

The detection of the five landmarks of a single patient was performed in two steps: a deep learning step and a post-processing step. During the deep learning step, the volumetric MDCT images were analysed by the three models and the output was combined and normalised to a final probability output volume that identified the regions of interest (Figure 6.5).

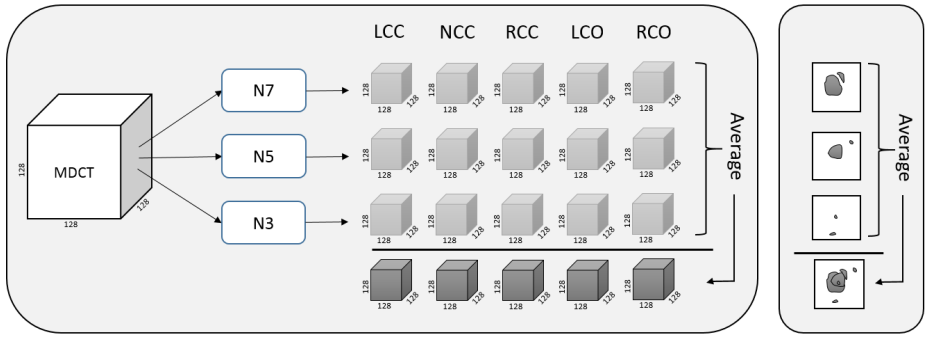



*Figure 6.5: Overview of the aortic landmark detection strategy in 3D (left) and for clarity in 2D (right). The MDCT volume is analysed by the three models each yielding five segmentation volumes. The output of the three models is averaged into five probability output volumes.*

During the post-processing step, all voxels with values above 0.5 were selected and clustered with hierarchical clustering (clustering details: Euclidean distance was used as criterion with a threshold of 1.1 mm). From the largest cluster, all points with values higher than 0.9 were used to compute the centroid, which was the final predicted point. This procedure was performed for each of the five landmarks (Figure 6.6). The output landmarks of the post-processing step were considered as the final predicted landmarks. After the detection phase, the aortic annular plane was reconstructed from the LCC, NCC and RCC. The distances from this plane to the LCO and RCO were measured. These measurements are referred to as the left- and right coronary ostium height.

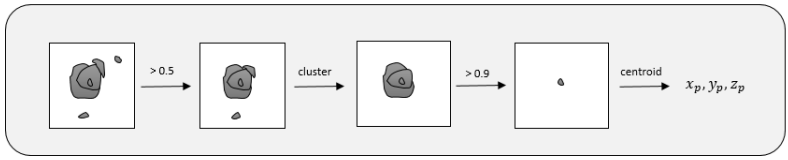



*Figure 6.6: Overview of the post-processing step depicted in 2D for clarity. From the averaged volume, the points with values higher than 0.5 were clustered. The largest cluster was selected and the centroid was computed from the points with values higher than 0.9. This centroid was used as the final predicted point.*

#### 6.2.2.4 Statistical Analysis

The accuracy of the landmark detection was assessed by measuring the Euclidean distances from the predicted landmarks to the ground truth values. The Shapiro-Wilk test was performed to test for normal distribution, and none of the predicted distributions was normally distributed. All variables were reported as median [lower quartile (LQ) - upper quartile (UQ)]. The agreement between manual and the automatic landmark locations were evaluated using the non-parametric signed Wilcoxon test (with a significant p-value $< 0.05$). Pearson correlation coefficient was computed for the coronary distances to observe their correlation (with excellent correlation $R^2$ $>0.9$)). The Bland-Altman analysis was performed for the coronary height results.

### 6.2.3 Results

The proposed method was validated using the 100 patients also used in the inter-observer variability study. By using the same patients for validation and observer variability assessment, it was possible to compare the method with both observers. The predicted landmarks were compared to the ground truth landmarks (observer one) using the Euclidean distance. The median difference for all landmarks combined was 1.5 mm, which is accurate considering the 1 mm resolution of the MDCT images used to obtain these predictions. The median Euclidean distance for all landmarks between the manually detected landmarks from the first and second observer was 2.0 mm. The distances between model and observer one are lower than the distances

between both observers, and the overall paired difference of -0.5 ± 1.3 mm is expected because the model was trained with data from observer one (Table 6.3).

| Landmark | Model vs. ground truth (d1) [mm] | Observer 1 vs. observer 2 (d2) [mm] | Paired diff. (d1-d2) | p-value |
|---|---|---|---|---|
| LCC | 1.6 [1.2-2.3] | 2.4 [1.4-3.4] | $0.8 \pm 1.3$ | <0.001 |
| NCC | 1.5 [0.9-2.1] | 2.4 [1.4-3.2] | $-0.9 \pm 1.3$ | <0.001 |
| RCC | 1.6 [1.3-2.2] | 2.4 [1.9-3.6] | $-0.9 \pm 1.4$ | <0.001 |
| LCO | 1.3 [0.9-1.9] | 1.4 [1.0-2.2] | $-0.1 \pm 1.0$ | 0.2 |
| RCO | 1.4 [0.9-2.0] | 1.4 [1.0-1.8] | $0.1 \pm 1.0$ | 0.4 |
| All | 1.5 [1.1-2.1] | 2.0 [1.3-2.8] | $-0.5 \pm 1.3$ | <0.001 |

*Table 6.3: Comparison of the Euclidean distances between the predicted and ground truth target points (d1), and the target points identified by the two observers (d2). Paired difference (Paired diff.) is in mm.*

As a final step, the left- and right coronary ostium heights were computed from ground truth and predicted points. There was a small overestimation by the model for the LCO height (model vs. observer one respectively: median height 16.3 mm vs. 15.8 mm) and a smaller overestimation for the RCO height (model vs. observer one respectively: median height 17.3 mm vs. 17.2 mm) (Table 6.4).

| Landmark | Corr. ($R^2$) | Model [mm] | Observer 1 [mm] | Paired diff. (model - observer 1) [mm] | p-value |
|---|---|---|---|---|---|
| LCO height | 0.8 | 16.3 [15.2-17.7] | 15.8 [14.7-17.5] | 0.5 ± 1.5 | <0.001 |
| RCO height | 0.8 | 17.3 [15.5-19.3] | 17.2 [15.1-20.1] | -0.2 ± 2.0 | 0.4 |

*Table 6.4: Comparison between the predicted (model) and the ground truth (observer 1) coronary ostium heights.*

The predicted left- and right coronary ostium height also correlated well with the ground truth (both with $R^2$=0.8). The left- and right coronary height derived from the manually detected landmarks of the second observer correlated well with the ground truth (respectively, $R^2 = 0.80$ and $R^2 = 0.84$) (Figure 6.7).

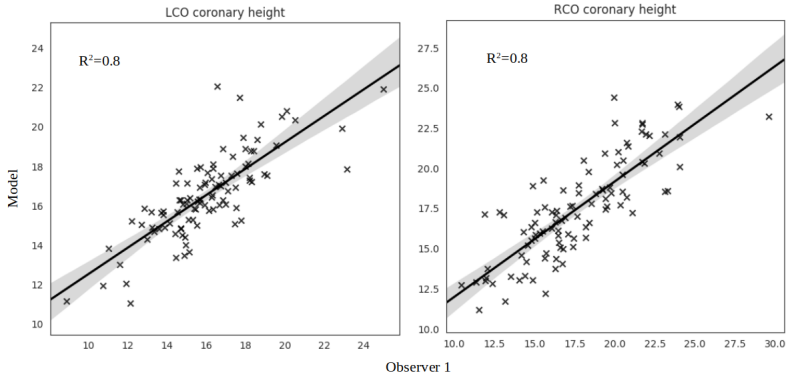



*Figure 6.7: The correlation between the predicted coronary ostium heights and the ground truth from observer 1.*

Bland-Altman plots for left- and right coronary ostium height were created. The LCO height presented with a mean paired difference of 0.54 mm for observer one whereas the RCO height resulted in a mean paired difference of -0.16 mm (Figure 6.8).

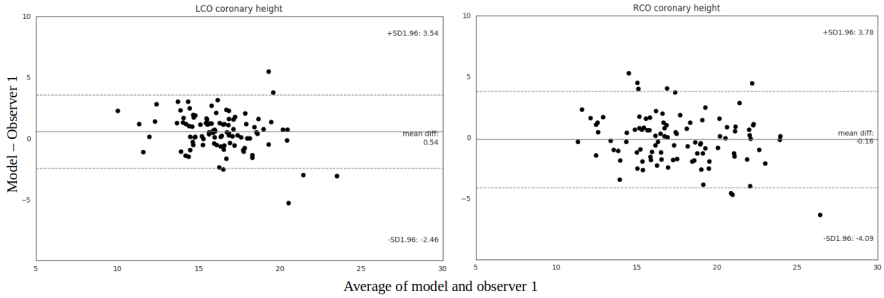



*Figure 6.8: Bland-Altman analysis for left- and right coronary ostium height.*

After the validation of the predicted landmarks and coronary heights, it remains to report the processing time for the manual and automated method. Both observers, reported 5 to 10 minutes of analysis time per patient to detect the five landmarks (the time to derive the coronary heights is ignored). The automatic processing time from volumetric MDCT image to the five predicted landmarks and the coronary heights is below 1 second. An example of the predicted output is depicted in Figure 6.9.

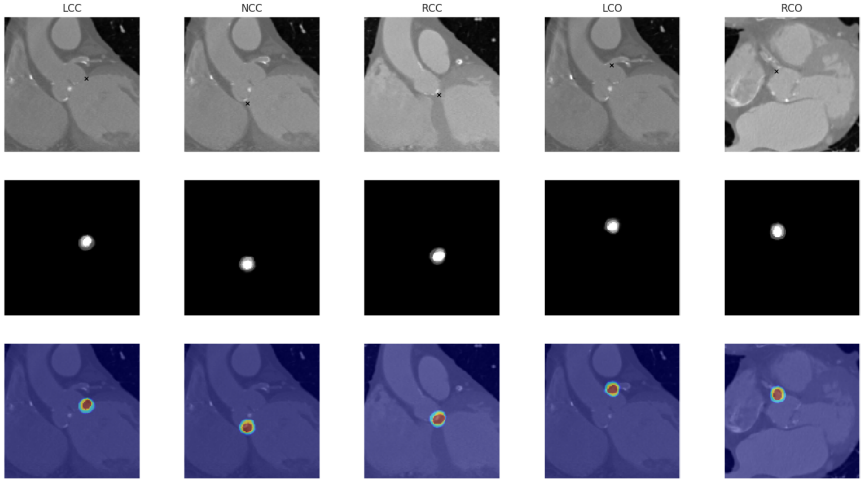



*Figure 6.9: An example of the detected landmarks from an unseen patient. From left to right: LCC, NCC, RCC, LCO and RCO. From top to bottom: a slice of the MDCT image at the location of the predicted point (annotated), a slice of the averaged final output volume at the location of the predicted point and an overlay of the final output volume on the MDCT image at the location of the predicted point (probability 0.0 (blue) to 1.0 (red)).*

### 6.2.4 Discussion

The literature offers (semi-) automatic strategies for landmarks detection in the TAVR field [209–212]. The limitations of these works can be summarised as follows: all studies validated their proposed method with single-center data, which leaves the question about the generic inter-center property of the method unanswered. Some automatic methods were still operator dependent. Therefore, the robustness of the method can not be properly assessed, and quality may depend on the experience of the operator. Some studies presented difficulties to adapt to specific pathological conditions, e.g. high calcium load, which is a common aspect for patients requiring a TAVR procedure. The size of the validation cohort was in most studies limited, and the inter-observer validation was in few studies absent or not entirely blind.

The presented method overcomes almost all aforementioned limitations. The methodology is fully automated and is insensitive to the amount of calcium load. Even if the patient cohort is relatively small, the number of patients used for training and validation is higher than in the reported literature. We used 344 patients for training and 100 for validation where we proved robustness of the method with a multi-centered patient cohort which displayed a good agreement with both observers.

Deep learning methods for landmark detection from medical images can be found in the literature. In Zheng *et al.*, a two-stage classification system for detecting landmarks from head-neck CT scans is proposed. The reported error is $2.6 \pm 5.0$ mm [200]. In Payer *et al.*, a method to detect landmarks from 2D CT-scans and 3D MR images using heatmaps is proposed with a reported error of $1.2 \pm 1.3$ mm [205]. In Zhang *et al.*, CNNs are used to detect landmarks of the brain from MR images with a reported error of $3.0 \pm 1.6$ mm. The same method is applied to detect landmarks of the prostate from CT-scans with a reported error of $3.3 \pm 2.5$ mm [189]. In Al *et al.*, a colonial walk method is proposed to detect similar aortic anatomical landmarks included in this study. Although the method is fast (12 ms per patient), the overall landmark error for non-TAVR patients is $1.94 \pm 0.93$ mm and for TAVR patients is $2.74 \pm 1.78$ mm [201]. In Lalys *et al.*, an automatic segmentation method is proposed. Their results show high precision for the two ostia with distance errors of $1.80 \pm 0.74$ mm and $1.96 \pm 0.87$ mm for the LCO and RCO. These errors are slightly higher compared to our results. The study does not include the hinge points of the aortic valve [213]. Finally, in O'Neil *et al.*, a two-pass method for localising 22 anatomical landmarks from head CT scans is proposed. Their method combines a neural network with a landmark atlas technique. The reported median error of 1.5 mm is equal to our findings. However, 10.8% of the detected landmarks had an error distance greater than 4.0 mm (computed on 20 scans and a total of 417 landmarks) [202]. Our results showed a 0.03% landmark error distance greater than 4.0 mm (computed on 100 scans and a total of 500 landmarks). Only in Payer *et al.*, the reported error is sufficiently fine-grained, which can be expected considering that MR images contain more details than CT images and bone landmarks are very recognisable landmarks in medical images. In our study, we obtained similar or more accurate results when detecting soft tissue landmarks in a coarser-grained image modality (CT). Since the reported

errors were too coarse-grained or the image modalities of the studies mentioned above differed from ours, a heatmap detection method with varying sizes is used.

The potential impact of this work may manifest itself in different frontiers. The method is faster than the state-of-the-art, which may have an impact on reducing operator analysis time in a rapidly growing market. Reduced overall TAVR costs may be obtained by embedding the method in software that allows manual corrections (e.g. to correct outliers). This embedding could also yield a continuous learning platform where the coordinates of a new patient, validated by an expert, can be added to the training dataset, thus improving future detections. The coronary ostia heights were derived from the predicted landmarks, which are vital measurements during the pre-operative planning of a TAVR procedure [214]. Other clinical parameters, such as the virtual basal ring's minimum and maximum diameter with the circumference, could be derived as well, which could be done by training a CNN to learn to predict these clinical parameters from images or with the method described in chapter 5. The same method could be applied for the planning of other cardiovascular interventions, e.g. left atrial appendage occlusion or mitral valve repair/replacement.

Although the presented method has proven to be reliable, there are a few limitations related to the current approach. The maximum outlier for the predicted LCC compared to the first observer was 4.9 mm, which is acceptable when compared to the inter-observer difference of 5.8 mm for the same patient. The difference between both observers might indicate a wrong detection by observer one. The maximum outlier of the predicted RCC was 11.5 mm, with an inter-observer measurement of 9.4 mm of the same patient. Investigation showed that the images of that patient contained movement, thus making the detection of the landmark difficult for the observer and the algorithm. The predicted measurement lies closer to the measurement of observer two, which may indicate that the models are capable of learning the general location of a landmark and thus contradicting the ground truth. The predicted NCC, LCO and RCO landmarks presented with maximum outliers smaller than 5 mm. The maximum outliers of the predicted coronary ostium heights were smaller than 5 mm. Although the predictive power has been shown, the outliers indicate that the method is not fail-safe. The final predicted result needs to be validated by an experienced operator, which could benefit the continuous learning method, as described in the previous section. As a final limitation, the MDCT images were of good to excellent quality with a few exceptions of images containing movement or having poor quality. Including poor quality MDCT scans, in the training dataset could improve the robustness of the method.

**Conclusion** In this work, an automated method is presented to extract landmarks for pre-operative TAVR planning from MDCT using deep learning techniques. When comparing to other methods in the available literature, the accuracy, efficiency of the method for detecting landmarks in soft tissue and blood environments has not been surpassed. Embedding this tool based on deep learning in the pre-operative planning routine may have an impact on the TAVR environments by reducing the time and cost and improving accuracy.

## 6.3 Curriculum deep reinforcement learning with different exploration strategies: A feasibility study on cardiac landmark detection

Although TAVR is less-invasive than SAVR, it is associated with an increased incidence of higher grade atrioventricular (AV) block that requires permanent pacemaker implantation [13]. The conduction abnormalities are due to the mechanical interaction between the prosthesis and the AV conduction path located near the aortic valve. Pre-operative assessment of the location of the AV conduction path and the expected contact pressure in this region can help to understand the risk of post-TAVR conduction abnormalities [215, 216]. As the AV conduction path itself is not visible on the cardiac CT images, the inferior border of the membranous septum can be used as an anatomical landmark (Figure 6.10). Detecting the inferior border of the membranous septum can be a difficult task and is dependent on operator experience, image quality, and the patient's anatomy. Detecting these points in an automatic, accurate, and efficient manner would save operator time and thus benefit pre-operative planning.

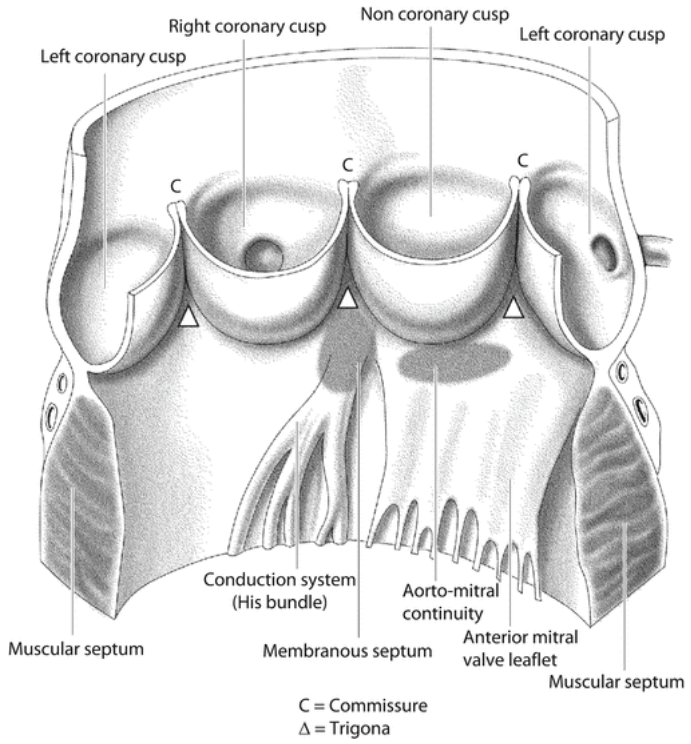



*Figure 6.10: Schematic drawing of the aortic root and the membranous septum. Image from [27].*

According to the literature, supervised deep learning methods are available that detect landmarks in medical images [189, 200–202, 205, 208]. These methods analyse a delimited region of interest and thus share the same limitation because they require a region of interest annotation. In this study, we wished to overcome this limitation and analyse medical images that range from scans of the entire body to region-specific scans.

Recent studies have shown that deep reinforcement learning can be used to detect landmarks in medical images regardless of their dimensions [204, 217]. In both studies, accurate results were obtained using a cluster of computational resources. However, the following difficulties may arise when exploring their methods. Deep reinforcement learning introduces novel hyper-parameters on top of the already existing hyper-parameters specific to deep learning and machine learning. When a cluster of computational resources is unavailable, the hyper-parameter search may require a considerable amount of time. Since deep reinforcement learning is still developing, it is essential to look into mechanisms that require additional research, e.g. the studies as mentioned earlier have different opinions on the definition of the halting signal for the artificial agent. Moreover, novel strategies should be researched and evaluated in a reproducible manner.

In this preliminary study, we aimed to identify the feasibility of 3D landmark detection in cardiac CT images with curriculum deep q-learning. This novel method of learning allowed us to use limited computational resources and tune the hyper-parameters accordingly. It also enabled us to evaluate three strategies that drive the exploring character of the artificial agent: the already established $\epsilon$-decay strategy and two novel strategies $\alpha$-decay and action-dropout. Each strategy was validated with k-folding; a cross-validation method that allows one to use the entire dataset for validation while maintaining a strict separation between the training and validation dataset. An inter-observer variability study was conducted to assess the accuracy of the method.

## 6.3.1 Method

### 6.3.1.1 Patient and imaging data

This retrospective study used the anonymised data of 278 patients from multiple centers. The average age of this cohort was 80.7 $\pm$ 6.6 years, and 56% of the patients were female. The patient data consisted of volumetric cardiac CT images, which were acquired to plan a TAVR procedure. Therefore, all CT images were contrast-enhanced and contained a certain degree of aortic stenosis. The dimensions of the images ranged from regional scans, centred around the aortic valve, to scans of the entire body. The volumetric CT images were resampled to obtain a homogeneous dataset and each image was interpolated to isotropic resolutions of 5.0, 3.0 2.0 and 1.0 mm using cubic spline interpolation. These different resolutions were used to enlarge the agent's field of view. A pre-processing step normalised all the images.

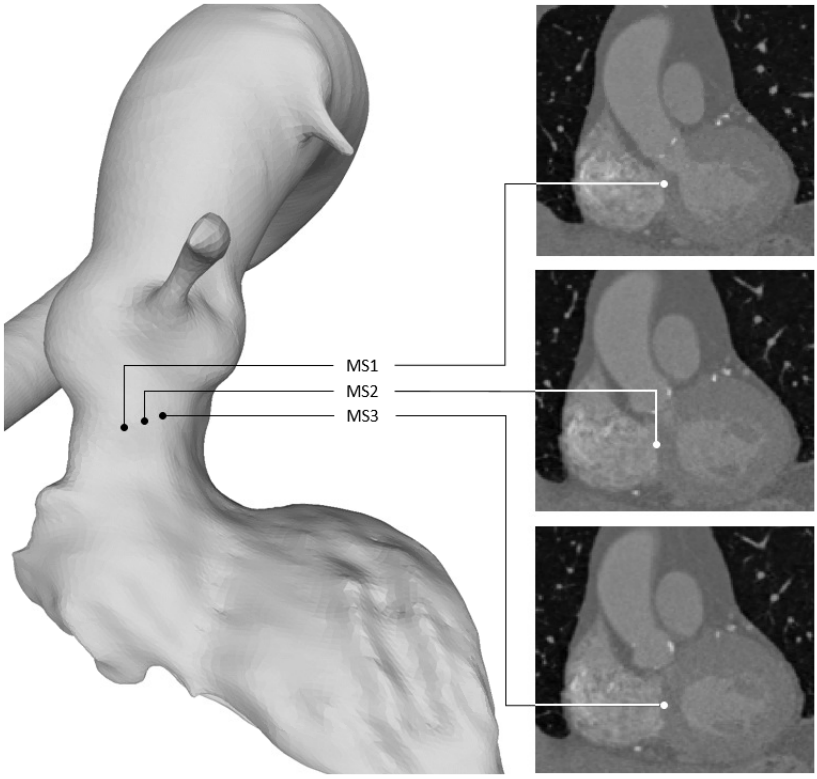



*Figure 6.11: The 3D view (left) and the coronal view (right) of the aortic valve with the annotated ground truth points: MS1, MS2 and MS3.*

An expert analysed all images and identified the inferior border of the membranous septum. This border was defined by three 3D points: MS1, MS2 and MS3 (Figure 6.11). These three landmarks were considered the ground truth in this study.

#### 6.3.1.2 Landmark detection with reinforcement learning

The task of detection landmarks in medical images with reinforcement learning is defined as follows. An artificial agent is positioned in an environment (an image) where it can walk around in search of the target landmark. Based on its current position, the agent has a limited field of view, causing it to see only a small image patch (state). From its current state **s**, the agent needs to choose an action **a**, which will move him closer to the target landmark. After performing that action, the agent will be at a new position in the image and thus have a new state **s'**. The reward **r** for choosing an action is the delta between the Euclidean distance from the previous- and current position to the target landmark. The agent's experience is defined as **s**, **a**, **s'** and **r** and needs to be stored in order to learn from past experiences. Thus, eventually improving the agent's decision-making in the future (Figure 6.12).

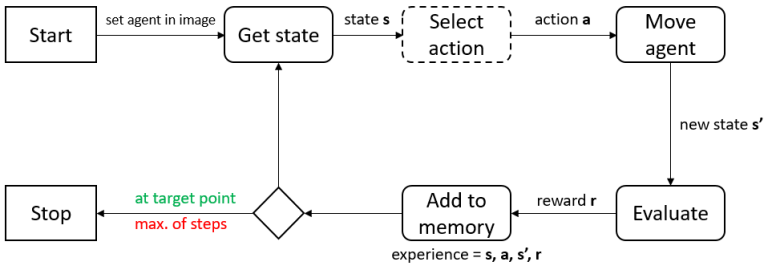



*Figure 6.12: Overview of landmark detection in a medical image using reinforcement learning. The agent will stop after a maximum number of steps or when it has found the target landmark. The 'select action' box in the figure is annotated with a dashed line because this is where the action-selection strategies will occur.*

#### 6.3.1.3 Landmark detection with deep reinforcement learning

The agent needs to have a map from states to actions to select the action that would yield the maximum future reward. Mapping all possible image patches in cardiac CT to actions is currently in terms of computational memory, impossible. Deep q-learning [218] therefore uses deep CNN to approach the required mapping (or q-function) from states to actions. In both studies [204, 217], a deep CNN or deep q-network (DQN) was trained with the experience of the agent to predict q-values. The size of the q-values coincided with the size of the action-space. The index of the highest q-value is selected as the action that yields the maximum future reward for the agent.

In this study, the agent was positioned in a cardiac CT image with multiple resolutions (5.0, 3.0, 2.0 and 1.0 mm) and all the movements of the agent were performed in the CT image with resolution 1.0 mm. The state **s** was defined by an 8x8x8 cube. In literature, state sizes of 45x45x45 (Alansary *et al.*) and 25x25x25 (Ghesu *et al.*) were selected however since multiple resolutions were used as input in our study, a state size of 8x8x8 sufficed. The agent could choose from 26 actions. This action-space covered all the possible single-step directions in a 3D grid (besides standing still). After each move, the experience of the agent was stored in prioritised experience replay memory [219]. This memory enabled the optimisation process to train with experiences which had a higher priority (or learning value) and resulted in faster convergence.

A double duelling DQN method was used to approximate the mapping between the states and actions. Two identical models were used to reduce the overestimation of the q-values. While the first DQN model was used to retrieve the actions of the current observations, a second (frozen) DQN model was used to yield q-values during the optimisation process [220]. After a variable number of iterations ($\tau$), the weights of the first DQN were transferred to the frozen DQN. In order to learn which states were valuable regardless of the effect of the action, the architectural modification proposed in [221] was used. The channel dimension of the model stores the multiple resolutions of the state. (Figure 6.13).

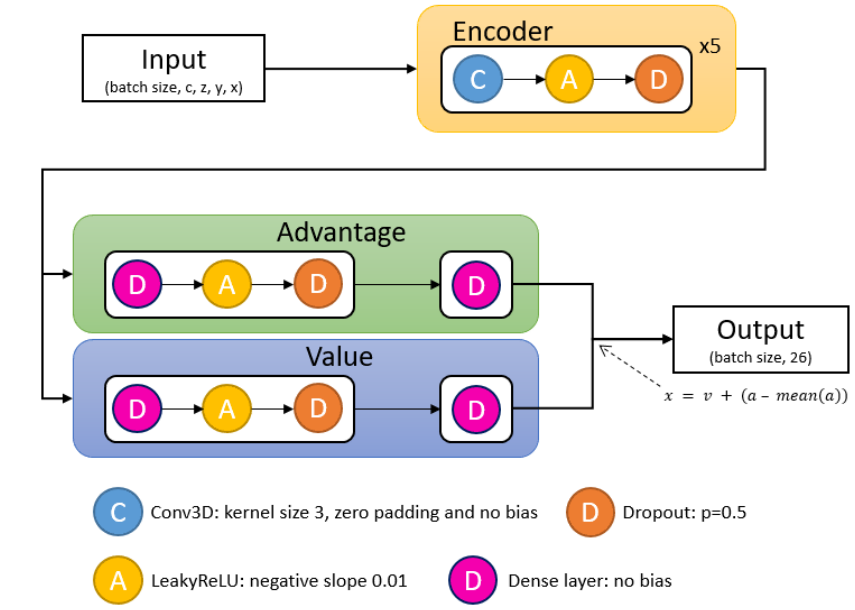



*Figure 6.13: The architecture of the duelling DQN model. The input has the dimensions of the state and the output of the model is the combination of the advantage- and value output signals [221].*

#### 6.3.1.4 Landmark detection with curriculum deep q-learning

Curriculum learning [222] is inspired by the human education system, where students start with a straightforward concept before learning more advanced concepts. This multiple-concept approach was already applied to object detection, where straightforward networks were trained first [82]. In this study, we defined the concepts as the distances to the target point T. The goal was to find a route from starting point S to T. The distance between these points was split into ten sub-starting points. The first and most straightforward concept was defined as the route from the closest sub-starting point $s^1$ to T. When the agent had learned this path, the subsequent concepts were processed until the model had learned the most difficult concept or route from S to T (Figure 6.14).

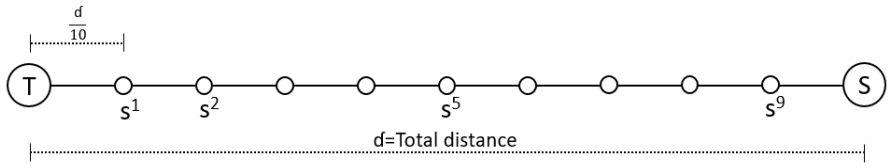



*Figure 6.14: Graphical representation of the curriculum learning approach. The final goal was to learn an entire pathway from the starting point S to the target point T.*

#### 6.3.1.5 The exploring agent

To learn the pathway towards the target point, the agent needs to explore and exploit paths. The exploration is required at the beginning of the training to discover new directions towards the target point. The exploitation is required near the end of the training to establish discovered routes. In this study, we compared three strategies to balance exploration and exploitation: the known exponential $\epsilon$-decay and two novel strategies: $\alpha$-decay and action-dropout.

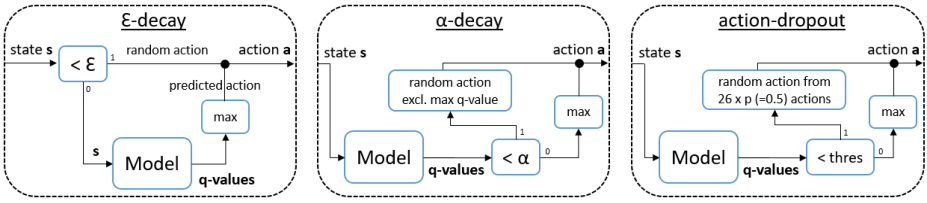



*Figure 6.15: Overview of the three action-selection strategies: $\epsilon$-decay, $\alpha$-decay and action-dropout.*

Exponential $\epsilon$-decay uses a value $\epsilon$ that drives the decision for predicting the action or randomly selecting an action. During training, an $\epsilon$ value needs to be decayed, and when a random value is smaller than the current $\epsilon$ value, random action is chosen. Otherwise, the model predicts the q-values from the current state and the action with the highest q-value is selected as the predicted action (Figure 6.15).

Linear $\alpha$-decay always uses the model to predict the q-values from the current state of the agent. An $\alpha$ value needs to be linearly decayed from 1.0 to 0.0 during training. If a random value is smaller than the current $\alpha$ value, the action with the highest q-value is ignored, and a random action between the remaining 25 actions is selected. When applied, this strategy ignores the highest q-value on purpose, causing the agent to find other routes to the target landmark. Otherwise, the action with the highest q-value is selected as the predicted action (Figure 6.15).

Similar to $\alpha$-decay, action-dropout uses the model to predict the q-values from the current state of the agent. A threshold value needs to be linearly decayed from 1.0 to 0.0 during training. Each time action-dropout is applied, a percentage $p$ of actions are removed from the action-space, and a random action is selected from the reduced action-space. If action-dropout is not applied, the action with the highest q-value is selected (Figure 6.15).

#### 6.3.1.6 Algorithmic details

The pseudo-code of the used curriculum deep q-learning algorithm is presented in Algorithm 1. The model learns the paths from starting point $s^1$ to $s^{10}$ or S. In this section; we explain a few parameters. The maximum number of steps that the agent could make was initially set to 15 to avoid the agent getting lost at the beginning of the algorithm. The algorithm gradually increased the maximum number of steps

to ensure that the agent had enough time to reach the target point. The number of iterations $\tau$ triggered the weight transfer from the first DQN to the frozen DQN and was initially set to a small value. Therefore, numerous patients would pass before the weight transfer. We observed that transferring the weights too quickly resulted in models which were not able to find the target point. The algorithm increased the parameter $\tau$ when the starting points were set further away from the target points.

**Algorithm 1** Pseudo code of the curriculum deep q-learning algorithm.

```
procedure TRAIN(patients)
    batchsize = 128
    model = initModel()
    model_frozen = model.copy()
    memory = initMemory()
    for i in [1..10] do
        τ = i * 500
        total number of steps = 0
        max steps = i * 15
        while threshold > minimum do
            p = getRandomPatient(patients)
            t = p.getTargetPoint()
            s^i = p.getStartingPoint(i)
            for step in max steps do
                state = p.getState()
                action = p.selectAction()
                reward, new state, is done = p.step(action)
                memory.push(state, action, new state, reward)
                optimize()
                total number of steps+ = 1
                if is done then
                    break
                end if
            end for
            if total number of steps % τ == 0 then
                model_frozen = model.copy()
            end if
        end while
    end for
end procedure
```

#### 6.3.1.7 Validation

The cohort size was relatively small (278 patients) and k-folding was employed as a cross-validation strategy to maximise the amount of data available for training and validating the models. K-folding (k=4) divided the entire dataset into four sub-sets, and the training and validation process was repeated four times or folds. Each fold used a different sub-set as the validation dataset, and the remaining three subsets were used as the training dataset. K-folding allowed us to use the entire dataset as the validation dataset while preserving the separation between training and validation dataset. Note that the validation process was similar to the training process except that the threshold value was set to -1. This value forced the agent to use the model to predict the action from the current state by selecting the action with the highest q-value from the predicted q-values.

### 6.3.2 Results

The validation process was applied to all three action-selection strategies. Cross-validation was applied to train and validate the method for all three target landmarks: MS1, MS2 and MS3. We obtained 278 validated patients per target landmark as a result of using k-folding. The action-selection strategies were compared to each other by grouping the results of the three target landmarks into 834 validated patients per strategy. A total of 24 trained models was used to validate one action-selection strategy.

The final path from starting point S to target point T was the longest and the most difficult one for the agent. The Euclidean distance from the agent's final position to the target point was measured. This final position was considered the predicted landmark of the method. The $\epsilon$-decay and action-dropout strategy yielded similar results. Whereas the $\alpha$-decay strategy was more accurate than the other strategies. The prediction time of the full path for a single patient was below 1 second (Table 6.5).

| **Name** | **Distance from target landmark [mm]** | | |
|---|---|---|---|
| | **$\epsilon$-decay** | **$\alpha$-decay** | **action-dropout** |
| MS1 | 2.4 [0.0 - 4.9] | 1.2 [0.0 - 4.2] | 2.3 [0.0 - 4.5] |
| MS2 | 3.0 [0.0 - 4.7] | 0.0 [0.0 - 4.6] | 2.0 [0.0 - 4.2] |
| MS3 | 2.2 [0.0 - 4.5] | 1.4 [0.0 - 4.3] | 1.0 [0.0 - 4.2] |
| All | 2.4 [0.0 - 4.6] | 0.0 [0.0 - 4.4] | 2.0 [0.0 - 4.4] |

*Table 6.5: A summary of the Euclidean distances from the predicted landmark to the target landmark for all three action-selection strategies.*

The agent learned the path from the starting point S to the target point T by gradually moving the starting point further away from T. The agent's performance during the curriculum learning process was measured by validating the models before moving to the next sub-starting point. This inter-algorithmic-validation resulted in Euclidean distances between the predicted points and T for each of the ten sub-starting points. In order to compare action-selection strategies, the distances for each starting point were grouped per strategy. The distances to the target landmarks reduced overtime for all three strategies. The $\alpha$-decay strategy is similar to the other strategies until the final starting point S, where the median distance to the target landmark was zero (Figure 6.16).

An inter-observer variability study was conducted on 20 patients to assess the accuracy of the method. The Euclidean distances between the landmarks annotated by the two observers were used as the difference between both observers. These differences were compared to the predicted landmarks of the $\alpha$-decay strategy of the same 20 patients. The differences between the predictions of the model and observer one were overall lower than the differences between observer one and two (Table 6.6).

| Name | $\alpha$-decay vs observer 1 ($d_1$) [mm] | Observer 1 vs observer 2 ($d_2$) [mm] | Paired diff. ($d_1 - d_2$) | p-value |
|---|---|---|---|---|
| MS1 | 2.7 [0.0 - 5.1] | 3.8 [2.8 - 5.6] | 0.2 ± 7.9 | 0.1 |
| MS2 | 0.0 [0.0 - 4.5] | 3.4 [1.8 - 4.4] | -1.3 ± 2.1 | 0.01 |
| MS3 | 3.0 [0.0 - 4.3] | 2.3 [1.4 - 4.8] | -0.1 ± 3.0 | 1.0 |
| All | 2.6 [0.0 - 4.5] | 3.1 [1.6 - 4.9] | -0.4 ± 5.1 | 0.4 |

*Table 6.6: Comparison of the Euclidean distances between the predicted and target landmarks ($d_1$), and the target points identified by the two observers ($d_2$). Paired difference (paired diff.) in mm.*

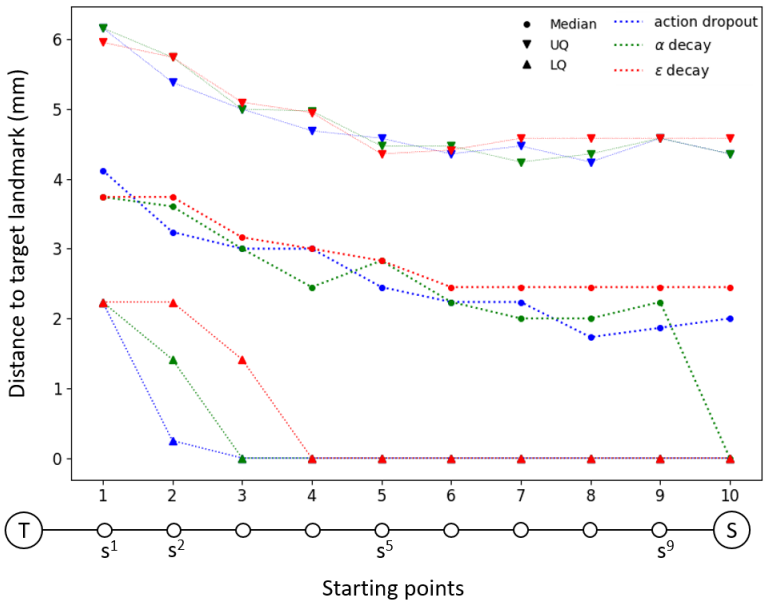



*Figure 6.16: The evolution of the distances to the target landmarks for each sub-starting point for all three action-selection strategies.*

### 6.3.3 Q-landscapes

The agent was trained to find a path from starting point S to target point T. During validation, the actions provided by the trained model guided the agent. In order to provide additional insight into the agent's world view, we plotted q-landscapes which show the highest q-values for the entire image. These landscapes were generated by convolving a state-sized cube over the entire image and selecting the highest q-value. This operation produced a volumetric q-landscape where each voxel represents the highest q-value of that particular position. In order to depict the 3D path of the agent, the path was projected on a plane. This plane was defined by the starting point, the target point and the centre of those two (Figure 6.17).

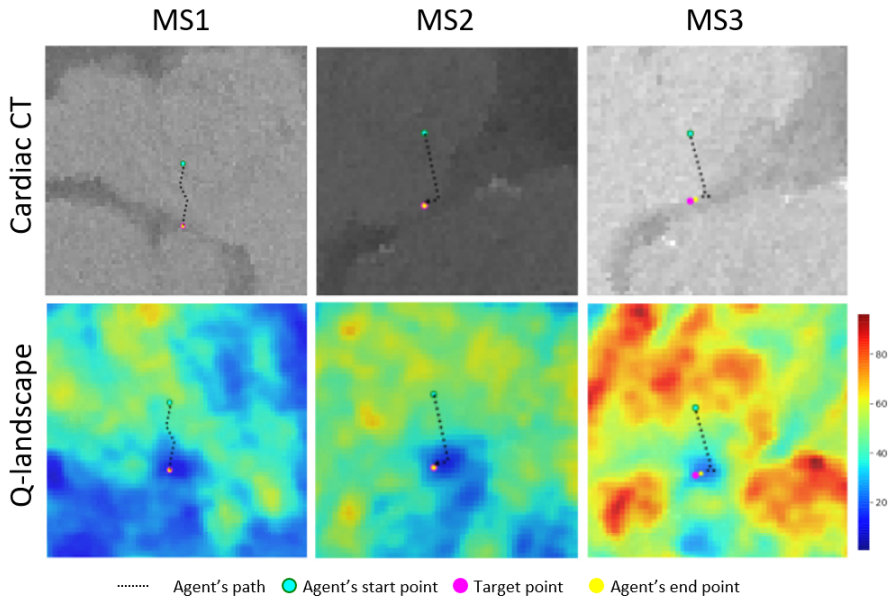



*Figure 6.17: The agent's path in the cardiac CT image (top) and the q-landscape (bottom) for each of the target landmarks: MS1, MS2 and MS3. The higher q-values are depicted in red and the lower q-values in blue. It is visible that the q-values around the target landmark are lower, as mentioned in [217].*

### 6.3.4 Discussion

In this feasibility study, a curriculum deep q-learning method was proposed to detect the inferior border of the membranous septum. We compared three action-selection strategies: $\epsilon$-decay, $\alpha$-decay and action-dropout and showed that $\alpha$-decay yielded the most accurate results. The method was efficient since the prediction time per patient was below 1 second, and the inter-observer variability study confirmed the accuracy of the method.

Previous studies [204, 217, 223] presented accurate and efficient landmark detection methods from medical images using deep q-learning. The authors used clusters of computational resources. In this study, the computational resources were limited to a single graphics processing unit (GPU). Additionally, these studies reported the usage of six actions (defined by the three Cartesian axes), whereas this study used 26 actions. This is to the best of our knowledge, the first time such a large action-space was used to detect landmarks from volumetric (medical) images. We expected that by using a larger action-space, convergence would take longer. However, we observed that convergence was faster when using 26 actions. By using curriculum learning, where easy tasks were learned first, the model was given the time and positive feedback to understand the task. It is our opinion that the incremental character of curriculum learning is responsible for this convergence speed-up.

Although this preliminary study has proven to be promising, there are a few

limitations to be noted. First, the model only learns the path from the centre of the aortic annular plane (with small random oscillations) to the target landmarks. When the agent enters areas outside the trained area or discovers image artefacts, it can get lost. This limitation will be handled in future work, wherein the agent will be put on the surface of a sphere around the target landmark with a gradually increasing radius. Next, the cohort size was relatively small and adding more patients will increase the robustness of the method. This limitation can also be addressed by applying data-augmentation.

There are a couple of recommendations for future work. First, a halting signal should be included. This mechanism can be obtained by adding a stopping action or measure oscillations in q-values as described by [204, 217]. Next, searching for multiple landmarks by using multiple agents could be considered [224]. Finally, the task of detecting landmarks with curriculum deep q-learning in volumetric images is computationally cheap yet difficult enough to serve as a baseline to experiment with novel deep reinforcement learning techniques.

**Conclusion** Curriculum deep q-learning was applied to detect the inferior border of the membranous septum. The output of this study is three-fold: first, the predicted landmarks were detected accurately and efficiently. Next, we have proposed and validated two novel action-selection strategies: $\alpha$-decay and action-dropout and compared it to the already established $\epsilon$-decay strategy. Finally, the incremental character of curriculum learning enabled the research into novel mechanisms and strategies in the domain of deep reinforcement learning.

## 6.4 Conclusion

In the first part of this chapter, we presented a literature study of landmark detection in medical images with deep learning methods between 2015 and 2019. In the second part of this chapter, two landmark detection methods are presented. The first study described an automatic method to detect five landmarks, which are crucial in the pre-operative phase of a TAVR procedure. After thorough validation, we concluded that the method was efficient and accurate and adopting this method in clinical practice may benefit the overall procedure in terms of time and accuracy. The second study uses curriculum DQL to detect the inferior border of the membranous septum. It was shown that this reinforcement learning approach could detect the target landmarks within acceptable accuracy.

# 7

# Conclusions and future perspectives

*We lose the precious sense*
*that an end*
*is only a beginning in disguise.*

Craig D. Lounsbrough

AI is rising fast and even surpassing human abilities on many levels. Furthermore, AI is already replacing human jobs. On the one hand, some scientists are tentative, telling us that AI can outperform humans, but note that very few systems are actually in place [225]. On the other hand, others are claiming that AI has achieved a medical degree, residency and fellowship degree in just one year, yet the evidence is lacking that AI may be used to bypass decision-making [226]. Others, try to look into the future by querying the experts of AI and conclude that there is a 50% chance that all human jobs will have been surpassed or replaced in 120 years [227]. These are all very different opinions, but the common denominator is that things will change, including the field of medical image analysis. How we are going to respond to this change is up to us. Baring in mind the AI winters of 1973 and 1987, where the hype could not fulfil its promises and funding would halt until the next AI spring [228], moving forward with care and caution may be the best thing to do.

The analysis of medical images is a field where human experts are essential; however, AI can contribute in various ways to the analysis. In this thesis, I have introduced and discussed studies where deep learning models captured expert knowledge and even went beyond human capabilities. For example, one significant study showed that retinal images images contained cardiovascular risk factors, which was thought not to be included in this image modality [89]. Such findings are astonishing and speak in favour of the analytical power of deep learning. Encapsulating such power in tools

that assist human experts can only benefit the analysis, which in return will benefit the patient.

Similar to deep learning, transcatheter treatments have gained considerable momentum in the past decade. Treatments, such as the transcatheter aortic valve replacement (TAVR), are established as a standard procedure while others, such as the transcatheter mitral valve replacement, still struggle with the technical and anatomical aspects due to the complexity of this valve.

This thesis focused on transcatheter heart valve interventions and the automation of some of the medical image analyses in this field. The number of patients and currently novel treatments will increase in the following years and since these interventions are closed-chest procedures, efficient and accurate medical image analysis will have a crucial influence on the patient's treatment and hospital time. The accuracy and efficiency of these presented methods provide the confidence for medical image analysis pipelines where images can be automatically analysed. This pipeline can be interpreted as a virtual expert or the physician's assistant, who has seen more patients than the physician yet is not flawless, and therefore has no role in the final decision-making process.

In this final chapter, we present an outline of the main conclusions from this work. Afterwards, some future directions are presented per study followed by some general future directions.

## 7.1 Research conclusions

In this thesis, cardiovascular medical image analysis methods, fueled by deep learning, have been presented. These methods can be categorised in classification, segmentation and landmark detection, which are all essential aspects of medical image analysis.

Classification plays a crucial role during medical image analysis. For example, the detection of bicuspid and tricuspid aortic valve patients influences the pre-operative planning of a TAVR. Segmentation is an essential medical image pre-processing step that assists in further critical anatomical measurements. Landmark detection influences crucial decisions and is vital for further analysis of the patient. For example, the proper detection of the three lowest hinge points of the aortic leaflets defines the aortic annular plane (AAP).

### 7.1.1 Classification

In chapter 4, we presented a retrospective study to classify aortic valve morphology from multidetector computed tomography (MDCT) imaging from patients who had been referred for transcatheter aortic valve replacement (TAVR). A three-dimensional (3D) convolutional neural network (CNN) was developed to classify the MDCT images and label the aortic valve morphology as tricuspid or bicuspid. The cross-validation of the dataset showed that a trained 3D CNN had an excellent ability to classify aortic valve morphology, with a high degree of both sensitivity (100%) and specificity

(92%). These results are very favourable, especially given the relatively small dataset on which the model was trained. Given these features, it is foreseeable that a tool such a this could be used by clinicians to identify BAV. When looking at the available literature, there are no studies which have used a 3D CNN model to classify entire 3D region of interest (ROI) as input.

### 7.1.2 Segmentation

In chapter 5, two studies have been presented where deep learning was applied to segment medical images in the transcatheter heart valve intervention domain. Both studies applied a post-processing step to derive clinically relevant measurements from the segmentation.

#### 7.1.2.1 Enabling automated device size selection for TAVR

In the first study, a method was proposed to facilitate and optimise the pre-operative TAVR planning by analysing the AAP from MDCT images. Measurements of the aortic annulus are vital during the decision-making process before a TAVR procedure. For example, the area or perimeter of the measured aortic annulus determine, among other measurements, the final prosthesis device size. The aortic annulus is measured on the AAP, a two-dimensional (2D) plane defined by the lowest hinge points of the aortic valve's leaflets.

In this retrospective study, the MDCT images of a substantial amount of patients were used to train two CNN models to segment the aortic annulus. The area and perimeter were derived from the predicted segmentation and were used to select prosthesis sizes for the Edwards Sapien 3 and the Medtronic Evolute device. The validation was performed in three steps. First, a test dataset of 118 patients was used to evaluate accuracy. These patients were excluded from the training phase of the models. Results showed that the area and perimeter could be predicted in an automatic, reproducible, efficient and accurate manner using our method. Secondly, the differences between the manually obtained aortic annulus measurements and the automatic predictions were similar to the differences between two independent observers. This indicates a satisfying accuracy of the proposed approach. Finally, when inserting the predicted area and perimeters in the sizing charts of the Edwards Sapien 3 and the Medtronic Evolute device, it was found that the automatically selected device sizes accorded well with the manually determined device sizes. The total analysis time from AAP to prosthesis size was below 1 second, which is much faster than the manual processing time of approximately 10 minutes. Therefore, our proposed method can speed-up the current pre-operative TAVR planning as well as reduce analysis errors in new centers.

When comparing this work to the available literature at the time of this study, many studies used deep learning to segment images. However, these studies analysed well-defined areas, e.g. the endocardium of the left ventricle (LV) can be delineated visually. In this study, the aortic annulus was segmented. The annular ring also delineates this structure, but calcified areas make the segmentation not straightforward.

The operator can include or exclude these calcified areas depending on the size of these areas. By segmenting the aortic annulus first, we could establish the confidence that further quantification would yield accurate results. We also expected that physicians would not appreciate black-box methods. If a model was trained to analyse the MDCT image and yield the area and perimeter as output directly, it would be difficult to convince a physician how the model came up with these values. Moreover, it would also be challenging to inspect the cause of outliers.

#### 7.1.2.2 Towards safe and efficient pre-operative planning of transcatheter mitral valve replacement (TMVR)

In the second study, a method was proposed to detect the mitral valve annulus from MDCT images automatically and derive clinically relevant measurements. The detection of the mitral valve annulus is a time-consuming and challenging process. The mitral valve annulus has a saddle-shaped anatomy and changes in size and shape during the cardiac cycle. From the detected mitral annulus, necessary measurements need to be assessed, such as the area, 2D and 3D perimeter, trigone-to-trigone-, septal-lateral- and commissure-to-commissure distance. The mitral valve annulus and the derived measurements are crucial during TMVR procedures and mitral valve prosthesis development.

This retrospective study used the MDCT images and expert data of a small number of patients to train deep learning models to segment the mitral valve annulus from medical images. Once segmented, the measurements described above were derived from the mitral valve annulus during a post-processing step. The method was cross-validated due to the small size of the dataset, which contained the image data of 71 patients. The accuracy of the method was confirmed by comparing the predicted and manually derived measurements. The total analysis time of our proposed method was less than 1 second per patient, which is much faster the manual processing time of approximately 25 minutes per patient. The efficiency and accuracy of our proposed method provide the confidence to move towards implementation of this technology in clinical practice. A proposal for a clinical application for medical image analysis was presented in this study.

In literature, there are currently no studies that used deep learning to segment a medical image to derive the mitral valve annulus and secondary measurements thereof. There are a few studies which use deep learning and other methods to detect the contours of the LV [130, 199]. The scope of this work targets a less-defined anatomical structure, i.e. the LV is well-delineated on CT images, whereas the mitral valve annulus is not.

### 7.1.3 Landmark detection

In chapter 6, two studies have been presented where deep learning was applied to detect landmarks from medical images.

#### 7.1.3.1 Aortic annular plane and coronary ostia detection

In the first study, an automated method was presented to extract landmarks for pre-operative TAVR planning from MDCT using deep learning techniques. During the pre-operative phase of a TAVR procedure, the MDCT images are used to identify the patient's risk. For example, the distances from the AAP to the coronary ostia help to understand the risk of coronary obstruction. This obstruction is a potentially life-threatening complication during which the blood flow to a coronary artery is significantly reduced. The AAP is defined by the three lowest hinge points of the aortic valve's leaflets: the left coronary cusp (LCC), the non coronary cusp (NCC) and the right coronary cusp (RCC).

In this retrospective study, the MDCT images and expert data of a substantial cohort were used to train models that could detect the five landmarks mentioned above. The method combined the results of three trained CNN models and a post-processing step. Three steps were applied to validate the method. First, a test dataset of 100 patients (these were excluded from the training phase) was used to evaluate the accuracy of our method and results showed that the five landmarks could be detected efficiently and accurately. Next, an inter-observer variability study was conducted. The difference between the manually and automatically identified landmarks was generally smaller when compared to differences observed between two independent operators. This indicates that the suggested approach detects these landmarks within acceptable accuracy. Finally, it was also illustrated that the method allows to accurately and automatically determine clinically relevant measurements such as coronary height. The total analysis time from MDCT image to the predicted landmarks was less than one second, which clearly shows the potential of our proposed method to speed up current pre-operative planning workflows.

It was shown that the accuracy and efficiency of the method for detecting landmarks in soft tissue environments have not been surpassed when compared to other methods in the available literature.

#### 7.1.3.2 Curriculum deep reinforcement learning with different exploration strategies: A feasibility study on cardiac landmark detection

In this second study, the curriculum deep q-learning (DQL) algorithm was used and applied to detect the inferior border of the membranous septum automatically. This border is a known anatomical landmark for the atrioventricular conduction path. Pre-operative assessment of the location of the atrioventricular conduction path is crucial to understand the risk of conduction abnormalities following a TAVR.

The idea behind curriculum DQL is that an artificial agent gradually learns to find the landmarks within the medical image. The agent has a limited view and learns to walk towards the target landmarks. During the walk, it tries to maximise some notion of cumulative reward. The model was cross-validated due to the small dataset, and the validation illustrated that the method was accurate and efficient. An inter-observer variability study confirmed the accuracy of the method. The output of this study

was three-fold: first, the predicted landmark locations were detected accurately and efficiently. Next, two novel action-selection strategies $\alpha$-decay and action-dropout were validated and compared to the already established $\epsilon$-decay. Finally, the task of detecting these landmarks was computationally cheap (when compared to similar methods) yet difficult enough to serve as a baseline to experiment with novel DQL techniques.

## 7.2 Future directions

In the following sections, we will present an outline of possible future directions for the presented studies in this thesis. This chapter concludes with a few general future directions.

### 7.2.1 Future directions per study

One future direction which is applicable for all the presented studies is the following: all studies would have benefited from more expert data. It should be apparent that including a wider variety of anatomical differences in the training dataset would make the trained models more resilient to anatomical differences. Moreover, the confidence in the validation of the proposed studies would increase by similarly expanding the test dataset. However, how much data is enough? The question remains how much data is required to perform a specific task. This question should be addressed in both ways. First, we can add more data to the dataset but this should enrich the variety of the dataset. For example, the patient's age. Preference should go to add patients with different ages in order to increase the variety in the dataset. Secondly, we execute performance measure studies where we remove data from the dataset in order to establish how much data is required to perform a task. This could be performed by using the current studies (and datasets) as baselines and incrementally reducing the amount of training data to perform the same task.

#### 7.2.1.1 Aortic valve morphology classification

In this study, a deep learning classification model was able to distinguish the bicuspid aortic valve patients from the tricuspid aortic valve patients (Section 4.2).

This study can be extended to a multi-class classification where all types of bicuspid aortic valves are included. Including all subcategories would provide more insights into the classification abilities of the deep learning model and may provide more clinical information to the physician. This method can be encapsulated in an automated pre-operative TAVR medical analysis tool. This tool would predict measurements and labels such as bicuspid aortic valve type from the medical images efficiently and accurately.

Both binary and multi-class studies would benefit from identifying the anatomical parts, which contributed to the final classification. This identification can be performed by using a classification activation map [229]. By localising the anatomical features that enabled correct predictions, we may gain novel insights to understand which

parts of the anatomy cause the classification. In its current status, this method can act as a virtual referee in case of doubt. However, incorporating the classification activation map may add signification value to the refereeing abilities.

The ability to classify a morphology or anatomical deformations from medical images applies to other fields. When limited to the field of cardiovascular medical image analysis, the following applications such as classifying the level of mitral annular calcification (MAC), atrial septal defect (ASD) or patent foramen ovale (PFO) from medical images comes to mind. Although severe MAC patients are visually easy to detect, capturing early-stage MAC patients may benefit the patient. Both ASD and PFO describe an unclosed foramen ovale located at the septum between the atria. This defect allows oxygen-rich blood to leak into the oxygen-poor blood chambers and may be automatically detectable with this architecture and strategy.

The 3D CNN classification model can be used as a baseline for architecture evolution research since this study yielded excellent accuracy. The 3D CNN model was developed by using state-of-the-art methods, parameter tuning methods and an enthusiastic dose of trial-and-error. Recent developments [230, 231] in deep learning architecture evolution have been made where architecture is evolved instead of developed. Small steps in this field were made, and preliminary results yielded surprising discoveries, i.e. the activation function is of less importance for simple tasks (see Appendix B). This 3D classification task is difficult enough to serve as a realistic benchmark to see if we can 1), outperform this architecture (with fewer parameters) and 2), potentially discover new efficient patterns of layers such as residual blocks or dense blocks with evolution.

#### 7.2.1.2 Enabling automated device size selection for TAVR

In this study, deep learning was used to automatically predict the TAVR device size from the aortic annular plane (Section 5.2).

The study was limited to the Edwards Lifesciences and Medtronic transcatheter aortic bioprostheses and could be extended to other devices. A different approach could be evaluated and compared to the current study: in order to circumvent the post-processing step, one could use a model with multiple outputs for segmentation, anatomical measurements and device sizes. If the results are sufficiently accurate, this multiple-output approach could yield an all-round aortic annular plane analysis tool.

Being able to measure area and perimeters in a 2D plane of the aortic annulus, enables the following research directions. Estimating the volume from the left ventricle (LV), right ventricle (RV), left atrium (LA) or right atrium (RA) by stacking the predicted segmentation. This has already been done, mainly for the LV and RV. Going a step further, one could measure the left ventricle outflow tract from medical images with the proposed strategy. This measurement is paramount in the field of mitral valve replacement [232]. In the field of left atrial appendage occlusion, an interesting measurement is the maximum diameter of the appendage. Based on this measurement, device sizes are determined in order to occlude the opening. The proposed method could be used to identify this measurement.

#### 7.2.1.3 Towards safe and efficient pre-operative planning of TMVR

In this study, the mitral valve annulus was segmented from 3D MDCT images and anatomical measurements were derived from this segmentation (Section 5.3). The manual detection of the mitral valve annulus is a time-consuming job, and experts say that it is more of an art than a science.

Since the mitral valve is a complex structure, more and more centers will move towards four-dimensional (4D) MDCT images in order to fully understand the valve's apparatus. A follow-up on this study would be to include multiple phases of the cardiac cycle, thus expanding the detection from 3D to 4D images. The same 3D architecture may perform this dimension extension, or when dealing with sequences of data, recurrent neural networks come to mind. It would be interesting to see how they could assist with this task, but a good starting point could be cardiac phase classification from 4D MDCT images. This initial step could be accomplished by adding recurrent layers inside the current 3D network as was done for 2D images in [166]. In Appendix C, a 4D frame classifier is presented which analyses 4D ultrasound images.

In this study, the ROI around the mitral valve annulus was segmented. Afterwards, the annulus was detected in a post-processing step. Similarly, the leaflets of the mitral valve could be approximated. Combined with the mitral annulus, this would provide a broader insight into the working of the entire mitral valve apparatus.

Going beyond the mitral valve field, other anatomical features which are difficult to detect can be measured - for example, the membranous septum area and anatomical elements of the tricuspid valve.

Another direction of this study would be to experiment with the detection of the mitral valve annulus as a 3D polygon directly. This can be accomplished by landmark detection as presented in Section 6.2, vector field prediction [130, 189], regression [163] or by combining image-based deep learning with geometric deep learning [233, 234]. This latter approach was, unfortunately, not included in this research due to time limitations.

#### 7.2.1.4 Aortic annular plane and coronary ostia detection

In this study, five landmarks were detected from MDCT images. These landmarks are crucial during the pre-operative planning of a TAVR procedure (Section 6.2).

The robustness of this method entails that 1), this method can also be encapsulated in an automatic pre-operative TAVR medical image analysis tool and 2), that this landmark detection strategy could be applied to many other problems. Many of the studies in this thesis start from the assumption that the images are centered around a ROI. This method could detect this center and thus serve as an automatic pre-processing tool for other automated methods.

Other applications for robust landmark detection may include mitral valve annulus detection, membranous septum detection, papillary muscle detection and other cardiovascular landmark detection problems. Other applications may include plane detection from medical images. In order to provide clinical feedback, known views

such as the four-chamber view can be detected using this method. A four-chamber view is a plane that is defined by the apex, mitral valve center and aortic valve center, which could be detected using the proposed strategy.

#### 7.2.1.5 Curriculum deep reinforcement learning with different exploration strategies: A feasibility study on cardiac landmark detection

In this study, three landmarks were located using deep reinforcement learning. An immediate follow-up study originates from a current limitation. The current algorithm starts at the center of the aortic root and tries to find the three landmarks. Generalising this would imply that the starting point should be placed anywhere in the medical image.

The current state of this study can also serve as a baseline for future experiments with DQL. This 3D environment and current landmark detection task are difficult enough to be challenging and inexpensive in terms of computation time. The following novel methods could be tested. First, different reward strategies could be implemented; for example, an agent receives a higher reward for finding a solution faster. Next, implementing different halting methods such as adding a dedicated action as the halting mechanism.

### 7.2.2 General future direction

#### 7.2.2.1 Continuous learning

The presented strategies for automatic medical image analysis in chapters 4, 5 and 6 used the images of cohorts of patients. These cohorts were collected, anonymised and stored appropriately for further analysis. The described methods in this thesis used the images and expert data to train models. The environment in which this took place was under control, which implies that the detected hyper-parameters and obtained accurate results are specific to that cohort. In most cases, a model learns how to mimic this expert but preferably, a model learns to generalise beyond this expert. It is at this point that the model can correct the mistakes of this expert which was seen in Section 6.2.

In this context, continuous learning refers to continuously learning models by adding new samples in the dataset. Usually, models are trained and validated on a dataset. When a previously unseen sample is added to the training dataset, the models need to be re-trained, validated and hopefully obtain the same or better results. The life-cycle of hyper-parameter re-calibration, re-training and re-validation of the models has not been developed yet. This life-cycle is preferably automatic and presumably requires the necessary tuning. For example, with the expected declining age of TAVR patients, a difference in anatomical features can be expected. Another example might involve adding data from another hospital with a different contrast agent policy. One possibility is to redo the hyper-parameter tuning, training and validation but considering that more and more automated methods will be transformed into pipelines, it might be beneficial to initiate the continuous learning approach. This

would also introduce novel hyper-parameters such as balancing of the training and test dataset. This balance could be determined by age, or gender or other meta-features. Moreover, introducing a continuous learning system would initiate a step towards real (artificial) intelligence which implies the ability to adapt.

#### 7.2.2.2 Medical image pipelines

In section 5.3, a web-based medical image pipeline has been proposed for the automatic analysis of the mitral valve annulus in a clinical environment. The schematic overview of the proposal is slightly adapted and depicted in Figure 7.1. The proposed medical image pipeline can be generalised as follows: the hospital transfers the DICOM images of the patient to a web-based platform where the images are automatically analysed with deep learning methods. The measurements are derived within seconds, and after quality control, a report is generated containing all required information and sent back to the hospital. The output of the quality control and the medical images the patient is paired and stored in a database. Since a trained, human operator performs quality control, the output may contribute to the continuous learning strategy as a new sample in the dataset. The proposed pipeline from section 5.3 is applicable to the studies described in sections 4.2, 5.2 and 6.2.

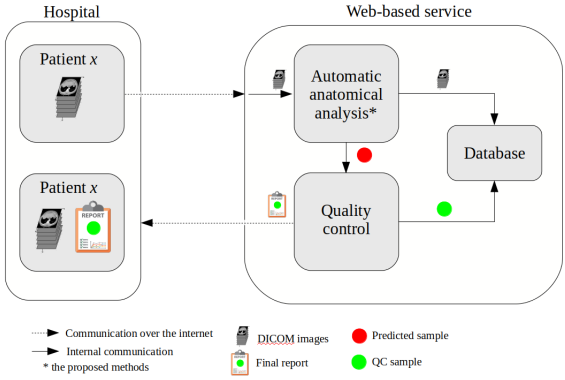



*Figure 7.1: Schematic overview of an generic web-based medical image analysis pipeline.*

#### 7.2.2.3 Learning with less data

As previously mentioned, all studies described in this thesis would benefit from more data, which means that more medical images need to be annotated by experts. In order to make claims that methods are accurate, preferably large varying datasets are required. In some of the presented studies in the field of TAVR, the dataset consisted of elderly patients with an average age of 80 years (Sections 6.2 and 5.2). However, since the age of TAVR patients is expected to decrease, and thus, younger patients will become candidates for the TAVR procedure, we can ask ourselves whether or not these models will be as accurate as they are now. One way to resolve this question is to train with more data. If we acquire more data, e.g. from younger patients, the

model should be resilient to varying anatomies. What if this data is not available or the number of older patients is much higher than the number of younger patients? Or the data is scarce because it is too expensive or painful to acquire?

One way to address this issue is to use less data, as was proposed in Madani *et al.* [88]. In this study, a semi-supervised method employed a generative adversarial network which used only 4% of the available expert data. The authors also used a supervised method that used all expert data. When comparing both methods, it was observed that both methods accomplished similar accurate results. This study showed that the hurdle of a small dataset could be resolved.

In the example of TAVR, where younger patients will need to be analysed, but the automatic methods were trained with older patients. If the dataset would be expanded with a few younger patients, it remains a question of which strategy would yield the most accurate results. One strategy could use the study mentioned above to tackle the dataset size or only increase the younger population in the dataset. Another strategy would be to re-balance the dataset where each age would be represented equally. Maybe less, properly balanced data is more.

# References


[1] F. Pesapane, M. Codari, and F. Sardanelli, "Artificial intelligence in medical imaging: Threat or opportunity? Radiologists again at the forefront of innovation in medicine", *European radiology experimental*, vol. 2, no. 1, p. 35, 2018.

[2] E. Coiera, "The fate of medicine in the time of Ai.", *Lancet (London, England)*, vol. 392, no. 10162, pp. 2331–2332, 2018.

[3] R. O. Bonow, B. A. Carabello, K. Chatterjee, *et al.*, "ACc/Aha 2006 guidelines for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing Committee to Revise the 1998 guidelines for the management of patients with valvular heart disease) developed in collaboration with the Society of Cardiovascular Anesthesiologists endorsed by the Society for Cardiovascular Angiography and Interventions and the Society of Thoracic Surgeons", *Journal of the American College of Cardiology*, vol. 48, no. 3, e1–e148, 2006.

[4] C. R. Smith, M. B. Leon, M. J. Mack, *et al.*, "Transcatheter versus surgical aortic-valve replacement in high-risk patients", *New England Journal of Medicine*, vol. 364, no. 23, pp. 2187–2198, 2011.

[5] L. Voigtländer and M. Seiffert, "Expanding TavI to low and intermediate risk patients", *Frontiers in cardiovascular medicine*, vol. 5, p. 92, 2018.

[6] G. Tarantini, L. Nai Fovino, and B. J. Gersh, "Transcatheter aortic valve implantation in lower-risk patients: What is the perspective?", *European heart journal*, vol. 39, no. 8, pp. 658–666, 2018.

[7] A. G. Cerillo, M. Mariani, M. Glauber, and S. Berti, "Sizing the annulus for transcatheter aortic valve implantation: More than a simple measure?", *European Journal of Cardio-Thoracic Surgery*, vol. 41, no. 3, pp. 717–718, 2012.

[8] A. Cribier, H. Eltchaninoff, A. Bash, *et al.*, "Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: First human case description", *Circulation*, vol. 106, no. 24, pp. 3006–3008, 2002.

[9] R. R. Makkar, G. P. Fontana, H. Jilaihawi, *et al.*, "Transcatheter aortic-valve replacement for inoperable severe aortic stenosis", *New England Journal of Medicine*, vol. 366, no. 18, pp. 1696–1704, 2012.

[10] J. J. Popma, D. H. Adams, M. J. Reardon, *et al.*, "Transcatheter aortic valve replacement using a self-expanding bioprosthesis in patients with severe aortic stenosis at extreme risk for surgery", *Journal of the American College of Cardiology*, vol. 63, no. 19, pp. 1972–1981, 2014.

[11] M. J. Mack, M. B. Leon, C. R. Smith, *et al.*, "5-year outcomes of transcatheter aortic valve replacement or surgical aortic valve replacement for high surgical risk patients with aortic stenosis (ParTneR 1): A randomised controlled trial", *The Lancet*, vol. 385, no. 9986, pp. 2477–2484, 2015.

[12] M. B. Leon, C. R. Smith, M. J. Mack, *et al.*, "Transcatheter or surgical aortic-valve replacement in intermediate-risk patients", *New England Journal of Medicine*, vol. 374, no. 17, pp. 1609–1620, 2016.

[13] M. J. Reardon, N. M. Van Mieghem, J. J. Popma, *et al.*, "Surgical or transcatheter aortic-valve replacement in intermediate-risk patients", *New England Journal of Medicine*, vol. 376, no. 14, pp. 1321–1331, 2017.

[14] J. J. Popma, G. M. Deeb, S. J. Yakubov, *et al.*, "Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients", *New England Journal of Medicine*, vol. 380, no. 18, pp. 1706–1715, 2019.

[15] M. J. Mack, M. B. Leon, V. H. Thourani, *et al.*, "Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients", *New England Journal of Medicine*, vol. 380, no. 18, pp. 1695–1705, 2019.

[16] V. T. Nkomo, J. M. Gardin, T. N. Skelton, J. S. Gottdiener, C. G. Scott, and M. Enriquez-Sarano, "Burden of valvular heart diseases: A population-based study", *The Lancet*, vol. 368, no. 9540, pp. 1005–1011, 2006.

[17] V. Bapat, V. Rajagopal, C. Meduri, *et al.*, "Early experience with new transcatheter mitral valve replacement", *Journal of the American College of Cardiology*, vol. 71, no. 1, pp. 12–21, 2018.

[18] G. W. Stone, J. Lindenfeld, W. T. Abraham, *et al.*, "Transcatheter mitral-valve repair in patients with heart failure", *New England Journal of Medicine*, vol. 379, no. 24, pp. 2307–2318, 2018.

[19] M. Guerrero, D. Dvir, D. Himbert, *et al.*, "Transcatheter mitral valve replacement in native mitral valve disease with severe mitral annular calcification: Results from the first multicenter global registry", *JACC: Cardiovascular Interventions*, vol. 9, no. 13, pp. 1361–1371, 2016.

[20] R. J. Schneider, N. A. Tenenholtz, D. P. Perrin, G. R. Marx, J. Pedro, and R. D. Howe, "Patient-specific mitral leaflet segmentation from 4D ultrasound", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2011, pp. 520–527.

[21] A. M. Pouch, C. Xu, P. A. Yushkevich, A. S. Jassar, M. Vergnat, J. H. Gorman III, R. C. Gorman, C. M. Sehgal, and B. M. Jackson, "Semi-automated mitral valve morphometry and computational stress analysis using 3D ultrasound", *Journal of biomechanics*, vol. 45, no. 5, pp. 903–907, 2012.

[22] R. Bellman, "The theory of dynamic programming", Rand corp santa monica ca, Tech. Rep., 1954.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[25] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation", in *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, 2016, pp. 565–571.

[26] E. Gibson, F. Giganti, Y. Hu, *et al.*, "Automatic multi-organ segmentation on abdominal Ct with dense v-networks", *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.

[27] M. Misfeld, E. I. Charitos, and H.-H. Sievers, "Acquired lesions of the aortic valve", in *Cardiac Surgery*, Springer, 2017, pp. 759–794.

[28] T. M. Mitchell *et al.*, *Machine learning*, 1997.

[29] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[31] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural networks*, vol. 61, pp. 85–117, 2015.

[32] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[33] I. B. A. TURING, "Computing machinery and intelligence-Am Turing", *Mind*, vol. 59, no. 236, p. 433, 1950.

[34] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[35] A. L. Samuel, "Some studies in machine learning using the game of checkers", *IBM Journal of research and development*, vol. 3, no. 3, pp. 210–229, 1959.

[36] H. J. Kelley, "Gradient theory of optimal flight paths", *Ars Journal*, vol. 30, no. 10, pp. 947–954, 1960.

[37] A. Ivakhnenko and V. G. Lapa, "Cybernetic predicting devices", Purdue Univ Lafayette Ind School of Electrical Engineering, Tech. Rep., 1966.

[38] A. G. Ivakhnenko, "Polynomial theory of complex systems", *IEEE transactions on Systems, Man, and Cybernetics*, no. 4, pp. 364–378, 1971.

[39] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[40] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[41] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[42] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net", *arXiv preprint arXiv: 1412.6806*, 2014.

[43] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.

[44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[45] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", *arXiv preprint arXiv: 1207.0580*, 2012.

[46] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks", *arXiv preprint arXiv: 1312.6120*, 2013.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv: 1409.1556*, 2014.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv: 1412.6980*, 2014.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv preprint arXiv: 1502.03167*, 2015.

[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[51] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions", *arXiv preprint arXiv: 1511.07122*, 2015.

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

[53] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[54] W. H. Organization *et al.*, "Hearts: Technical package for cardiovascular disease management in primary health care", 2016.

[55] J. P. Fanning, D. L. Walters, D. G. Platts, E. Eeles, J. Bellapart, and J. F. Fraser, "Characterization of neurological injury in transcatheter aortic valve implantation: How clear is the picture?", *Circulation*, vol. 129, no. 4, pp. 504–515, 2014.

[56] H. Jilaihawi, M. Kashif, G. Fontana, *et al.*, "Cross-sectional computed tomographic assessment improves accuracy of aortic annular sizing for transcatheter aortic valve replacement and reduces the incidence of paravalvular aortic regurgitation", *Journal of the American College of Cardiology*, vol. 59, no. 14, pp. 1275–1286, 2012.

[57] D. H. Adams, J. J. Popma, M. J. Reardon, *et al.*, "Transcatheter aortic-valve replacement with a self-expanding prosthesis", *New England Journal of Medicine*, vol. 370, no. 19, pp. 1790–1798, 2014.

[58] S. M. O'Brien, D. M. Shahian, G. Filardo, *et al.*, "The Society of Thoracic Surgeons 2008 cardiac surgery risk models: Part 2—isolated valve surgery", *The Annals of thoracic surgery*, vol. 88, no. 1, S23–S42, 2009.

[59] H. Baumgartner, V. Falk, J. J. Bax, *et al.*, "2017 Esc/EacTs guidelines for the management of valvular heart disease", *European heart journal*, vol. 38, no. 36, pp. 2739–2791, 2017.

[60] M. Enriquez-Sarano, V. T. Nkomo, and H. I. Michelena, "Mitral regurgitation", in *Valvular Heart Disease*, Springer, 2009, pp. 221–246.

[61] R. Hasan, V. S. Mahadevan, H. Schneider, and B. Clarke, "First in human transapical implantation of an inverted transcatheter aortic valve prosthesis to treat native mitral valve stenosis", *Circulation*, vol. 128, no. 6, e74–e76, 2013.

[62] J.-M. Sinning, F. Mellert, W. Schiller, A. Welz, G. Nickenig, and C. Hammerstingl, "Transcatheter mitral valve replacement using a balloon-expandable prosthesis in a patient with calcified native mitral valve stenosis", *European heart journal*, vol. 34, no. 33, pp. 2609–2609, 2013.

[63] N. M. Van Mieghem, N. Piazza, R. H. Anderson, *et al.*, "Anatomy of the mitral valvular complex and its implications for transcatheter interventions for mitral regurgitation", *Journal of the American College of Cardiology*, vol. 56, no. 8, pp. 617–626, 2010.

[64] C. P. Huded and M. Y. Desai, "Early experience with transcatheter mitral valve replacement: Successes, challenges, and future directions", *Journal of thoracic disease*, vol. 10, no. Suppl 9, S1008, 2018.

[65] A. Cheung, J. G. Webb, D. R. Wong, J. Ye, J.-B. Masson, R. G. Carere, and S. V. Lichtenstein, "Transapical transcatheter mitral valve-in-valve implantation in a human", *The Annals of Thoracic Surgery*, vol. 87, no. 3, e18–e20, 2009.

[66] A. W. Cheung, R. Gurvitch, J. Ye, D. Wood, S. V. Lichtenstein, C. Thompson, and J. G. Webb, "Transcatheter transapical mitral valve-in-valve implantations for a failed bioprosthesis: A case series", *The Journal of Thoracic and Cardiovascular Surgery*, vol. 141, no. 3, pp. 711–715, 2011.

[67] D. Dvir, A. Assali, H. Vaknin-Assa, A. Sagie, Y. Shapira, E. Porat, and R. Kornowski, "Transcatheter aortic and mitral valve implantations for failed bioprosthetic heart valves", *Journal of Invasive Cardiology*, vol. 23, no. 9, p. 377, 2011.

[68] M. Seiffert, L. Conradi, S. Baldus, J. Schirmer, M. Knap, S. Blankenberg, H. Reichenspurner, and H. Treede, "Transcatheter mitral valve-in-valve implantation in patients with degenerated bioprostheses", *JACC: Cardiovascular Interventions*, vol. 5, no. 3, pp. 341–349, 2012.

[69] A. Cheung, J. G. Webb, M. Barbanti, M. Freeman, R. K. Binder, C. Thompson, D. A. Wood, and J. Ye, "5-year experience with transcatheter transapical mitral valve-in-valve implantation for bioprosthetic valve dysfunction", *Journal of the American College of Cardiology*, vol. 61, no. 17, pp. 1759–1766, 2013.

[70] B. D. Prendergast, H. Baumgartner, V. Delgado, *et al.*, "Transcatheter heart valve interventions: Where are we? Where are we going?", *European heart journal*, vol. 40, no. 5, pp. 422–440, 2019.

[71] J. Schofer, K. Bijuklic, C. Tiburtius, L. Hansen, A. Groothuis, and R. T. Hahn, "First-in-human transcatheter tricuspid valve repair in a patient with severely regurgitant tricuspid valve", *Journal of the American College of Cardiology*, vol. 65, no. 12, pp. 1190–1195, 2015.

[72] A. Latib, A. Pozzoli, P. Denti, *et al.*, "First-in-man implantation of a tricuspid annular remodeling device for functional tricuspid regurgitation", *JACC: Cardiovascular Interventions*, vol. 8, no. 13, e211–e214, 2015.

[73] P. Bonhoeffer, Y. Boudjemline, Z. Saliba, *et al.*, "Percutaneous replacement of pulmonary valve in a right-ventricle to pulmonary-artery prosthetic conduit with valve dysfunction", *The Lancet*, vol. 356, no. 9239, pp. 1403–1405, 2000.

[74] A. Chatterjee, N. S. Bajaj, W. S. McMahon, M. G. Cribbs, J. S. White, A. Mukherjee, and M. A. Law, "Transcatheter pulmonary valve implantation: A comprehensive systematic review and meta-analyses of observational studies", *Journal of the American Heart Association*, vol. 6, no. 8, e006432, 2017.

[75] H. Caughron, D. Kim, N. Kamioka, *et al.*, "Repeat pulmonary valve replacement: Similar intermediate term outcomes with surgical and transcatheter procedures", *JACC: Cardiovascular Interventions*, vol. 11, no. 24, pp. 2495–2503, 2018.

[76] G. T. Herman, *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[77] M. Moradi, Y. Gur, H. Wang, P. Prasanna, and T. Syeda-Mahmood, "A hybrid learning approach for semantic labeling of cardiac Ct slices and recognition of body position", in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016, pp. 1418–1421.

[78] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Išgum, "2D image classification for 3D anatomy localization: Employing deep convolutional neural networks", in *Medical imaging 2016: Image processing*, International Society for Optics and Photonics, vol. 9784, 2016, 97841Y.

[79] M. Zreik, T. Leiner, B. D. De Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, "Automatic segmentation of the left ventricle in cardiac Ct angiography using convolutional neural networks", in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016, pp. 40–43.

[80] J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, "Automatic coronary artery calcium scoring in cardiac Ct angiography using paired convolutional neural networks", *Medical image analysis*, vol. 34, pp. 123–136, 2016.

[81] O. Emad, I. A. Yassine, and A. S. Fahmy, "Automatic localization of the left ventricle in cardiac MRI images using deep learning", in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 683–686.

[82] L. Zhang, A. Gooya, B. Dong, R. Hua, S. E. Petersen, P. Medrano-Gracia, and A. F. Frangi, "Automated quality assessment of cardiac Mr images using convolutional neural networks", in *International Workshop on Simulation and Synthesis in Medical Imaging*, Springer, 2016, pp. 138–145.

[83] M. A. Gülsün, G. Funka-Lea, P. Sharma, S. Rapaka, and Y. Zheng, "Coronary centerline extraction via optimal flow paths and Cnn path pruning", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 317–325.

[84] A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, "Quantification of coronary artery calcium using ultrafast computed tomography", *Journal of the American College of Cardiology*, vol. 15, no. 4, pp. 827–832, 1990.

[85] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan, "A deep convolutional neural network model to classify heartbeats", *Computers in biology and medicine*, vol. 89, pp. 389–396, 2017.

[86] F. Pereira, A. Bueno, A. Rodriguez, D. Perrin, G. Marx, M. Cardinale, I. Salgo, and P. del Nido, "Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms", *Journal of Medical Imaging*, vol. 4, no. 1, p. 014 502, 2017.

[87] G. Santini, D. Della Latta, N. Martini, G. Valvano, A. Gori, A. Ripoli, C. L. Susini, L. Landini, and D. Chiappino, "An automatic deep learning approach for coronary artery calcium segmentation", in *EMBEC & NBC 2017*, Springer, 2017, pp. 374–377.

[88] A. Madani, J. R. Ong, A. Tibrewal, and M. R. Mofrad, "Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease", *NPJ digital medicine*, vol. 1, no. 1, pp. 1–11, 2018.

[89] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning", *Nature Biomedical Engineering*, vol. 2, no. 3, p. 158, 2018.

[90] J. D. Dormer, M. Halicek, L. Ma, C. M. Reilly, E. Schreibmann, and B. Fei, "Convolutional neural networks for the detection of diseased hearts using Ct images and left atrium patches", in *Medical Imaging 2018: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 10575, 2018, p. 1 057 530.

[91] J. F. Silva, J. M. Silva, A. Guerra, S. Matos, and C. Costa, "Ejection fraction classification in transthoracic echocardiography using a deep learning approach", in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2018, pp. 123–128.

[92] F. Commandeur, M. Goeller, J. Betancur, *et al.*, "Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast Ct", *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1835–1846, 2018.

[93] K. C. Wong, T. Syeda-Mahmood, and M. Moradi, "Building medical image classifiers with very limited data using segmentation networks", *Medical image analysis*, vol. 49, pp. 105–116, 2018.

[94] R. Shadmi, V. Mazo, O. Bregman-Amitai, and E. Elnekave, "Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest Ct", in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 24–28.

[95] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning", *Medical image analysis*, vol. 43, pp. 54–65, 2018.

[96] K. H. Miao and J. H. Miao, "Coronary heart disease diagnosis using deep neural networks", *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, pp. 1–8, 2018.

[97] N. Lessmann, B. van Ginneken, M. Zreik, P. A. de Jong, B. D. de Vos, M. A. Viergever, and I. Išgum, "Automatic calcium scoring in low-dose chest Ct using deep neural networks with dilated convolutions", *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 615–625, 2017.

[98] C. Mazo, J. Bernal, M. Trujillo, and E. Alegre, "Transfer learning for classification of cardiovascular tissues in histological images", *Computer methods and programs in biomedicine*, vol. 165, pp. 69–76, 2018.

[99] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[100] N. Gessert, M. Lutz, M. Heyder, S. Latus, D. M. Leistner, Y. S. Abdelwahed, and A. Schlaefer, "Automatic plaque detection in IvoCt pullbacks using convolutional neural networks", *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 426–434, 2018.

[101] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, "Motion artifact recognition and quantification in coronary Ct angiography using convolutional neural networks", *Medical image analysis*, vol. 52, pp. 68–79, 2019.

[102] A. Østvik, E. Smistad, S. A. Aase, B. O. Haugen, and L. Lovstakken, "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks", *Ultrasound in medicine & biology*, vol. 45, no. 2, pp. 374–384, 2019.

[103] R. Togo, K. Hirata, O. Manabe, H. Ohira, I. Tsujino, K. Magota, T. Ogawa, M. Haseyama, and T. Shiga, "Cardiac sarcoidosis classification with deep convolutional neural network-based features using polar maps", *Computers in biology and medicine*, vol. 104, pp. 81–86, 2019.

[104] S. L. Oh, E. Y. Ng, R. San Tan, and U. R. Acharya, "Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types", *Computers in biology and medicine*, vol. 105, pp. 92–101, 2019.

[105] J. Dodge, S. Gururangan, D. Card, R. Schwartz, and N. A. Smith, "Show your work: Improved reporting of experimental results", *arXiv preprint arXiv:1909.03004*, 2019.

[106] W. C. Roberts, "The congenitally bicuspid aortic valve: A study of 85 autopsy cases", *The American journal of cardiology*, vol. 26, no. 1, pp. 72–83, 1970.

[107] J. Steinberger, J. H. Moller, J. M. Berry, and A. R. Sinaiko, "Echocardiographic diagnosis of heart disease in apparently healthy adolescents", *Pediatrics*, vol. 105, no. 4, pp. 815–818, 2000.

[108] S. C. Siu and C. K. Silversides, "Bicuspid aortic valve disease", *Journal of the American College of Cardiology*, vol. 55, no. 25, pp. 2789–2800, 2010.

[109] H.-H. Sievers and C. Schmidtke, "A classification system for the bicuspid aortic valve from 304 surgical specimens", *The Journal of thoracic and cardiovascular surgery*, vol. 133, no. 5, pp. 1226–1233, 2007.

[110] G. Y. Perlman, P. Blanke, D. Dvir, *et al.*, "Bicuspid aortic valve stenosis: Favorable early outcomes with a next-generation transcatheter heart valve in a multicenter study", *JACC: Cardiovascular Interventions*, vol. 9, no. 8, pp. 817–824, 2016.

[111] X. Liu, Y. He, Q. Zhu, F. Gao, W. He, L. Yu, Q. Zhou, M. Kong, and J. Wang, "Supra-annular structure assessment for self-expanding transcatheter heart valve size selection in patients with bicuspid aortic valve", *Catheterization and Cardiovascular Interventions*, vol. 91, no. 5, pp. 986–994, 2018.

[112] T.-Y. Xiong, Y. Feng, Y.-J. Li, Z.-G. Zhao, Y.-B. Liao, Y. Ou, X. Wei, and M. Chen, "Supra-annular sizing for transcatheter aortic valve replacement candidates with bicuspid aortic valve", *JACC: Cardiovascular Interventions*, vol. 11, no. 17, pp. 1789–1790, 2018.

[113] R. Jeger, O. Reuthebuch, G. Fahrni, J.-C. Laborde, R. Vogel, and C. Kaiser, "Supra-annular sizing for transcatheter valve implantation in bicuspid aortic stenosis", *Postępy w Kardiologii Interwencyjnej= Advances in Interventional Cardiology*, vol. 14, no. 2, p. 187, 2018.

[114] A. Masri, L. G. Svensson, B. P. Griffin, and M. Y. Desai, "Contemporary natural history of bicuspid aortic valve disease: A systematic review", *Heart*, vol. 103, no. 17, pp. 1323–1330, 2017.

[115] A. L. W. Shroyer, M. E. Plomondon, F. L. Grover, F. H. Edwards, *et al.*, "The 1996 coronary artery bypass risk model: The society of thoracic surgeons adult cardiac national database", *The Annals of thoracic surgery*, vol. 67, no. 4, pp. 1205–1208, 1999.

[116] L. Breiman, "Bagging predictors", *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[117] D. Mylotte, G. Martucci, and N. Piazza, "Patient selection for transcatheter aortic valve implantation: An interventional cardiology perspective", *Annals of cardiothoracic surgery*, vol. 1, no. 2, p. 206, 2012.

[118] S.-H. Yoon, Y. Maeno, H. Kawamori, *et al.*, "Diagnosis and outcomes of transcatheter aortic valve implantation in bicuspid aortic valve stenosis", *Interventional Cardiology Review*, vol. 13, no. 2, p. 62, 2018.

[119] A. Colombo and A. Latib, *Bicuspid aortic valve: Any room for TAVR?*, 2014.

[120] R. A. Guyton and M. Padala, *Transcatheter aortic valve replacement in bicuspid aortic stenosis: Early success but concerning red flags*, 2016.

[121] R. Tanaka, K. Yoshioka, H. Niinuma, S. Ohsawa, H. Okabayashi, and S. Ehara, "Diagnostic value of cardiac Ct in the evaluation of bicuspid aortic stenosis: Comparison with echocardiography and operative findings", *American Journal of Roentgenology*, vol. 195, no. 4, pp. 895–899, 2010.

[122] H. Alkadhi, S. Leschka, P. T. Trindade, G. Feuchtner, P. Stolzmann, A. Plass, and S. Baumueller, "Cardiac Ct for the differentiation of bicuspid and tricuspid aortic valves: Comparison with echocardiography and surgery", *American Journal of Roentgenology*, vol. 195, no. 4, pp. 900–908, 2010.

[123] S. Verma and S. C. Siu, "Aortic dilatation in patients with bicuspid aortic valve", *New England Journal of Medicine*, vol. 370, no. 20, pp. 1920–1929, 2014.

[124] P. J. van Rosendael, V. Kamperidis, W. K. Kong, A. R. van Rosendael, N. A. Marsan, J. J. Bax, and V. Delgado, "Comparison of quantity of calcific deposits by multidetector computed tomography in the aortic valve and coronary arteries", *The American journal of cardiology*, vol. 118, no. 10, pp. 1533–1538, 2016.

[125] G. Carneiro, J. C. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods", *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 968–982, 2011.

[126] G. Carneiro and J. C. Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data", *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2592–2607, 2013.

[127] T. A. Ngo and G. Carneiro, "Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks", in *2013 IEEE International Conference on Image Processing*, IEEE, 2013, pp. 695–699.

[128] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI", *Medical image analysis*, vol. 30, pp. 108–119, 2016.

[129] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI", *arXiv preprint arXiv: 1604.00494*, 2016.

[130] C. Rupprecht, E. Huaroc, M. Baust, and N. Navab, "Deep active contours", *arXiv preprint arXiv: 1607.05074*, 2016.

[131] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin, "Cardiac left ventricle segmentation using convolutional neural network regression", in *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, 2016, pp. 490–493.

[132] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, and S. K. Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 487–495.

[133] F. C. Ghesu, E. Krubasik, B. Georgescu, V. Singh, Y. Zheng, J. Hornegger, and D. Comaniciu, "Marginal space deep learning: Efficient architecture for volumetric image parsing", *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1217–1228, 2016.

[134] J. C. Nascimento and G. Carneiro, "Multi-atlas segmentation using manifold learning with deep belief networks", in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016, pp. 867–871.

[135] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation", *Medical image analysis*, vol. 30, pp. 120–129, 2016.

[136] X. Yang, L. Gobeawan, S. Y. Yeo, W. T. Tang, Z. Wu, and Y. Su, "Automatic segmentation of left ventricular myocardium by deep convolutional and de-convolutional neural networks", in *2016 Computing in Cardiology Conference (CinC)*, IEEE, 2016, pp. 81–84.

[137] S. Andermatt, S. Pezold, and P. Cattin, "Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data", in *Deep Learning and Data Labeling for Medical Applications*, Springer, 2016, pp. 142–151.

[138] S. Dong, G. Luo, G. Sun, K. Wang, and H. Zhang, "A combined multi-scale deep learning and random forests approach for direct left ventricular volumes estimation in 3D echocardiography", in *2016 Computing In Cardiology Conference (Cinc)*, IEEE, 2016, pp. 889–892.

[139] J. Lieman-Sifry, M. Le, F. Lau, S. Sall, and D. Golden, "FastVentricle: Cardiac segmentation with Enet", in *International Conference on Functional Imaging and Modeling of the Heart*, Springer, 2017, pp. 127–138.

[140] L. V. Romaguera, M. G. F. Costa, F. P. Romero, and C. F. F. Costa Filho, "Left ventricle segmentation in cardiac MRI images using fully convolutional neural networks", in *Medical Imaging 2017: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 10134, 2017, 101342Z.

[141] A. H. Curiale, F. D. Colavecchia, P. Kaluza, R. A. Isoardi, and G. Mato, "Automatic myocardial segmentation by using a deep learning network in cardiac MRI", in *2017 XLIII Latin American Computer Conference (CLEI)*, IEEE, 2017, pp. 1–6.

[142] A. Giannakidis, K. Kamnitsas, V. Spadotto, *et al.*, "Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3D convolutional neural network", in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE, 2016, pp. 42–46.

[143] J. Li, R. Zhang, L. Shi, and D. Wang, "Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3D FCN", in *Reconstruction, Segmentation, and Analysis of Medical Images*, Springer, 2016, pp. 111–118.

[144] G. Luo, G. Sun, K. Wang, S. Dong, and H. Zhang, "A novel left ventricular volumes prediction method based on deep learning network in cardiac MRI", in *2016 Computing in Cardiology Conference (CinC)*, IEEE, 2016, pp. 89–92.

[145] G. Luo, R. An, K. Wang, S. Dong, and H. Zhang, "A deep learning network for right ventricle segmentation in short-axis MRI", in *2016 Computing in Cardiology Conference (CinC)*, IEEE, 2016, pp. 485–488.

[146] L. Yu, X. Yang, J. Qin, and P.-A. Heng, "3D fractalNet: Dense volumetric segmentation for cardiovascular MRI volumes", in *Reconstruction, segmentation, and analysis of medical images*, Springer, 2016, pp. 103–110.

[147] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images", *Medical image analysis*, vol. 41, pp. 40–54, 2017.

[148] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease", in *Reconstruction, segmentation, and analysis of medical images*, Springer, 2016, pp. 95–102.

[149] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin, "Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine Mr sequences", *Medical image analysis*, vol. 39, pp. 78–86, 2017.

[150] F. Liao, X. Chen, X. Hu, and S. Song, "Estimation of the volume of the left ventricle from MRI images using deep neural networks", *IEEE transactions on cybernetics*, vol. 49, no. 2, pp. 495–504, 2017.

[151] X. Yang, Z. Zeng, and S. Yi, "Deep convolutional neural networks for automatic segmentation of left ventricle cavity from cardiac magnetic resonance images", *IET Computer Vision*, vol. 11, no. 8, pp. 643–649, 2017.

[152] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach", *Magnetic resonance in medicine*, vol. 78, no. 6, pp. 2439–2448, 2017.

[153] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance", *Medical image analysis*, vol. 35, pp. 159–171, 2017.

[154] A. Kabani and M. R. El-Sakka, "Ejection fraction estimation using a wide convolutional neural network", in *International Conference Image Analysis and Recognition*, Springer, 2017, pp. 87–96.

[155] X. Yang, C. Bian, L. Yu, D. Ni, and P.-A. Heng, "Hybrid loss guided convolutional networks for whole heart parsing", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 215–223.

[156] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 111–119.

[157] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D u-Net: Learning dense volumetric segmentation from sparse annotation", in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 424–432.

[158] L. V. Romaguera, F. P. Romero, C. F. F. Costa Filho, and M. G. F. Costa, "Myocardial segmentation in cardiac magnetic resonance images using fully convolutional neural networks", *Biomedical Signal Processing and Control*, vol. 44, pp. 48–57, 2018.

[159] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Multi-label whole heart segmentation using CNNs and anatomical label configurations", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 190–198.

[160] S. Dong, G. Luo, K. Wang, S. Cao, Q. Li, and H. Zhang, "A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography", *BioMed research international*, vol. 2018, 2018.

[161] O. Oktay, E. Ferrante, K. Kamnitsas, *et al.*, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation", *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2017.

[162] L. Zhang, G. V. Karanikolas, M. Akçakaya, and G. B. Giannakis, "Fully automatic segmentation of the right ventricle via multi-task deep neural networks", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 6677–6681.

[163] L. K. Tan, R. A. McLaughlin, E. Lim, Y. F. Abdul Aziz, and Y. M. Liew, "Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression", *Journal of Magnetic Resonance Imaging*, vol. 48, no. 1, pp. 140–152, 2018.

[164] X. Du, W. Zhang, H. Zhang, J. Chen, Y. Zhang, J. C. Warrington, G. Brahm, and S. Li, "Deep regression segmentation for cardiac bi-ventricle Mr images", *IEEE Access*, vol. 6, pp. 3828–3838, 2018.

[165] Q. Zheng, H. Delingette, N. Duchateau, and N. Ayache, "3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation", *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2137–2148, 2018.

[166] N. Savioli, M. S. Vieira, P. Lamata, and G. Montana, "Automated segmentation on the entire cardiac cycle using a deep learning work-flow", in *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, IEEE, 2018, pp. 153–158.

[167] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features", in *International workshop on statistical atlases and computational models of the heart*, Springer, 2017, pp. 120–129.

[168] D. M. Vigneault, W. Xie, C. Y. Ho, D. A. Bluemke, and J. A. Noble, "Ω-Net (Omega-Net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks", *Medical image analysis*, vol. 48, pp. 95–106, 2018.

[169] J. Duan, G. Bello, J. Schlemper, *et al.*, "Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach", *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2151–2164, 2019.

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

[171] A. Mortazi, J. Burt, and U. Bagci, "Multi-planar deep segmentation networks for cardiac substructures from MRI and CT", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 199–206.

[172] C. Zotti, Z. Luo, O. Humbert, A. Lalande, and P.-M. Jodoin, "GridNet with automatic shape prior registration for automatic MRI cardiac segmentation", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 73–81.

[173] J. Patravali, S. Jain, and S. Chilamkurthy, "2D-3D fully convolutional neural networks for cardiac Mr segmentation", in *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, 2017, pp. 130–139.

[174] H. B. Winther, C. Hundt, B. Schmidt, C. Czerner, J. Bauersachs, F. Wacker, and J. Vogel-Claussen, "$\nu$-net: Deep Learning for Generalized Biventricular Mass and Function Parameters Using Multicenter Cardiac MRI Data", *JACC: Cardiovascular Imaging*, p. 2479, 2018.

[175] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling", *arXiv preprint arXiv: 1412.3555*, 2014.

[176] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers", *Medical image analysis*, vol. 51, pp. 21–45, 2019.

[177] A. H. Curiale, F. D. Colavecchia, and G. Mato, "Automatic quantification of the LV function and mass: A deep learning approach for cardiovascular MRI", *Computer methods and programs in biomedicine*, vol. 169, pp. 37–50, 2019.

[178] C. Petitjean, M. A. Zuluaga, W. Bai, *et al.*, "Right ventricle segmentation from cardiac mri: A collation study", *Medical image analysis*, vol. 19, no. 1, pp. 187–202, 2015.

[179] Y. Du, Y. Zhu, S. Wu, W. Lihui, Z. YueMin, and Y. Feng, "Automatic Segmentation of Left Myocardium in CMR Based on Fully Convolutional Networks", in *Proceedings of the 3rd International Conference on Crowd Science and Engineering*, 2018, pp. 1–4.

[180] A. Mangieri, C. Montalto, M. Pagnesi, G. Lanzillo, O. Demir, L. Testa, A. Colombo, and A. Latib, "TAVI and post procedural cardiac conduction abnormalities", *Frontiers in cardiovascular medicine*, vol. 5, p. 85, 2018.

[181] A. P. Durko, R. L. Osnabrugge, N. M. Van Mieghem, M. Milojevic, D. Mylotte, V. T. Nkomo, and A. Pieter Kappetein, "Annual number of candidates for transcatheter aortic valve implantation per country: Current estimates and future projections", *European heart journal*, vol. 39, no. 28, pp. 2635–2642, 2018.

[182] A. M. Kasel, S. Cassese, S. Bleiziffer, M. Amaki, R. T. Hahn, A. Kastrati, and P. P. Sengupta, "Standardized imaging for aortic annular sizing: Implications for transcatheter valve selection", *JACC: Cardiovascular Imaging*, vol. 6, no. 2, pp. 249–262, 2013.

[183] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks", in *Advances in neural information processing systems*, 2017, pp. 971–980.

[184] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations", in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2017, pp. 240–248.

[185] J. Canny, "A computational approach to edge detection", *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[186] S. Queirós, C. Dubois, P. Morais, T. Adriaenssens, J. C. Fonseca, J. L. Vilaça, and J. D'hooge, "Automatic 3D aortic annulus sizing by computed tomography in the planning of transcatheter aortic valve implantation", *Journal of cardiovascular computed tomography*, vol. 11, no. 1, pp. 25–32, 2017.

[187] S. Queirós, A. Papachristidis, D. Barbosa, K. C. Theodoropoulos, J. C. Fonseca, M. J. Monaghan, J. L. Vilaca, and J. D'hooge, "Aortic valve tract segmentation from 3d-tee using shape-based b-spline explicit active surfaces", *IEEE transactions on medical imaging*, vol. 35, no. 9, pp. 2015–2025, 2016.

[188] S. Queirós, A. Papachristidis, P. Morais, K. C. Theodoropoulos, J. C. Fonseca, M. J. Monaghan, J. L. Vilaça, and J. D'hooge, "Fully automatic 3-d-tee segmentation for the planning of transcatheter aortic valve implantation", *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1711–1720, 2016.

[189] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks", *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4753–4764, 2017.

[190] D. D. Wang, M. H. Eng, A. B. Greenbaum, *et al.*, "Validating a prediction modeling tool for left ventricular outflow tract (LVOT) obstruction after transcatheter mitral valve replacement (TMVR)", *Catheterization and Cardiovascular Interventions*, vol. 92, no. 2, pp. 379–387, 2018.

[191] P. Blanke, C. Naoum, J. Webb, *et al.*, "Multimodality imaging in the context of transcatheter mitral valve replacement: Establishing consensus among modalities and disciplines", *JACC: Cardiovascular Imaging*, vol. 8, no. 10, pp. 1191–1208, 2015.

[192] J. Karády, I. Ntalas, B. Prendergast, C. Blauth, S. Niederer, P. Maurovich-Horvat, and R. Rajani, "Transcatheter mitral valve replacement in mitral annulus calcification–"The art of computer simulation"", *Journal of cardiovascular computed tomography*, vol. 12, no. 2, pp. 153–157, 2018.

[193] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-D medial surface axis thinning algorithms", *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.

[194] K. Paton, "An algorithm for finding a fundamental set of cycles of a graph", *Communications of the ACM*, vol. 12, no. 9, pp. 514–518, 1969.

[195] I. Voigt, M. Scutaru, T. Mansi, B. Georgescu, N. El-Zehiry, H. Houle, and D. Comaniciu, "Robust live tracking of mitral valve annulus for minimally-invasive intervention guidance", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 439–446.

[196] M. van Stralen, K. Leung, M. M. Voormolen, N. de Jong, A. F. van der Steen, J. H. Reiber, and J. G. Bosch, "Time continuous detection of the left ventricular long axis and the mitral valve plane in 3-D echocardiography", *Ultrasound in medicine & biology*, vol. 34, no. 2, pp. 196–207, 2008.

[197] J. Pedrosa, S. Queirós, J. Vilaça, L. Badano, and J. D'hooge, "Fully automatic assessment of mitral valve morphology from 3D transthoracic echocardiography", in *2018 IEEE International Ultrasonics Symposium (IUS)*, IEEE, 2018, pp. 1–6.

[198] G. J. Mak, P. Blanke, K. Ong, *et al.*, "Three-dimensional echocardiography compared with computed tomography to determine mitral annulus size before transcatheter mitral valve implantation", *Circulation: Cardiovascular Imaging*, vol. 9, no. 6, e004176, 2016.

[199] J. C. Nascimento and G. Carneiro, "Deep learning on sparse manifolds for faster object segmentation", *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4978–4990, 2017.

[200] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu, "3D deep learning for efficient and robust landmark detection in volumetric data", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 565–572.

[201] W. A. Al, H. Y. Jung, I. D. Yun, Y. Jang, H.-B. Park, and H.-J. Chang, "Automatic aortic valve landmark localization in coronary Ct angiography using colonial walk", *PloS one*, vol. 13, no. 7, 2018.

[202] A. Q. O'Neil, A. Kascenas, J. Henry, *et al.*, "Attaining human-level performance with atlas location autocontext for anatomical landmark detection in 3D cT data", in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[203] J. M. Noothout, B. D. de Vos, J. M. Wolterink, T. Leiner, and I. Išgum, "CNn-based Landmark Detection in Cardiac Cta Scans", *arXiv preprint arXiv: 1804.04963*, 2018.

[204] F.-C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Multi-scale deep reinforcement learning for real-time 3D-landmark detection in Ct scans", *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 1, pp. 176–189, 2017.

[205] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Regressing heatmaps for multiple landmark localization using Cnns", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 230–238.

[206] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning", *IEEE transactions on medical imaging*, vol. 36, no. 10, pp. 2057–2067, 2017.

[207] S. Dangi, C. A. Linte, and Z. Yaniv, "Cine cardiac MRI slice misalignment correction towards full 3D left ventricle segmentation", in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, International Society for Optics and Photonics, vol. 10576, 2018, p. 1 057 607.

[208] Y. Zheng, M. John, R. Liao, A. Nottling, J. Boese, J. Kempfert, T. Walther, G. Brockmann, and D. Comaniciu, "Automatic aorta segmentation and valve landmark detection in C-arm Ct for transcatheter aortic valve implantation", *IEEE transactions on medical imaging*, vol. 31, no. 12, pp. 2307–2321, 2012.

[209] V. Delgado, A. C. Ng, J. D. Schuijf, *et al.*, "Automated assessment of the aortic root dimensions with multidetector row computed tomography", *The Annals of thoracic surgery*, vol. 91, no. 3, pp. 716–723, 2011.

[210] Y. Watanabe, M.-C. Morice, E. Bouvier, *et al.*, "Automated 3-dimensional aortic annular assessment by multidetector computed tomography in transcatheter aortic valve implantation", *JACC: Cardiovascular Interventions*, vol. 6, no. 9, pp. 955–964, 2013.

[211] A. Van Linden, J. Kempfert, J. Blumenstein, H. Möllmann, W.-K. Kim, S. Alkaya, C. Hamm, and T. Walther, "Manual versus automatic detection of aortic annulus plane in a computed tomography scan for transcatheter aortic valve implantation screening", *European Journal of Cardio-Thoracic Surgery*, vol. 46, no. 2, pp. 207–212, 2014.

[212] M. Elattar, E. Wiegerinck, F. van Kesteren, L. Dubois, N. Planken, E. Vanbavel, J. Baan, and H. Marquering, "Automatic aortic root landmark detection in Cta images for preprocedural planning of transcatheter aortic valve implantation", *The international journal of cardiovascular imaging*, vol. 32, no. 3, pp. 501–511, 2016.

[213] F. Lalys, S. Esneault, M. Castro, L. Royer, P. Haigron, V. Auffret, and J. Tomasi, "Automatic aortic root segmentation and anatomical landmarks detection for TavI procedure planning", *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 3, pp. 157–164, 2019.

[214] J.-B. Masson, J. Kovac, G. Schuler, *et al.*, "Transcatheter aortic valve implantation: Review of the nature, management, and avoidance of procedural complications", *JACC: Cardiovascular Interventions*, vol. 2, no. 9, pp. 811–820, 2009.

[215] A. Hamdan, V. Guetta, R. Klempfner, *et al.*, "Inverse relationship between membranous septal length and the risk of atrioventricular block in patients undergoing transcatheter aortic valve implantation", *JACC: Cardiovascular Interventions*, vol. 8, no. 9, pp. 1218–1228, 2015.

[216] G. Rocatello, N. El Faquir, G. De Santis, *et al.*, "Patient-specific computer simulation to elucidate the role of contact pressure in the development of new conduction abnormalities after catheter-based implantation of a self-expanding aortic valve", *Circulation: Cardiovascular Interventions*, vol. 11, no. 2, e005344, 2018.

[217] A. Alansary, O. Oktay, Y. Li, *et al.*, "Evaluating reinforcement learning agents for anatomical landmark detection", *Medical image analysis*, vol. 53, pp. 156–164, 2019.

[218] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, "Human-level control through deep reinforcement learning", *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[219] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay", *arXiv preprint arXiv: 1511.05952*, 2015.

[220] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning", in *Thirtieth AAAI conference on artificial intelligence*, 2016.

[221] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning", *arXiv preprint arXiv: 1511.06581*, 2015.

[222] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning", in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[223] F. C. Ghesu, B. Georgescu, T. Mansi, D. Neumann, J. Hornegger, and D. Comaniciu, "An artificial agent for anatomical landmark detection in medical images", in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 229–237.

[224] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning", in *International conference on machine learning*, 2016, pp. 1928–1937.

[225] L. The, "Artificial intelligence in health care: Within touching distance.", *Lancet (London, England)*, vol. 390, no. 10114, p. 2739, 2018.

[226] D. A. Bluemke, "Radiology in 2018: Are you working with Ai or being replaced by Ai?", *Radiology*, vol. 287, no. 2, pp. 365–366, 2018.

[227] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "When will Ai exceed human performance? Evidence from Ai experts", *Journal of Artificial Intelligence Research*, vol. 62, pp. 729–754, 2018.

[228] W. S. Bainbridge, "AI: The Tumultuous History of the Search for Artificial Intelligence", *Science*, vol. 261, no. 5125, pp. 1186–1187, 1993.

[229] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.

[230] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers", in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, JMLR. org, 2017, pp. 2902–2911.

[231] L. Xie and A. Yuille, "Genetic cnn", in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1379–1388.

[232] P. Blanke, C. Naoum, D. Dvir, *et al.*, "Predicting LVOT obstruction in transcatheter mitral valve implantation: Concept of the neo-LVOT", *JACC: Cardiovascular Imaging*, vol. 10, no. 4, pp. 482–485, 2017.

[233] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model", *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.

[234] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5115–5124.

# A

# Towards realistic and tailored medical image generation

## A.1 Introduction

Data-sciences such as deep learning require large datasets. Bigger datasets are generally speaking equivalent to more robust models. When data is expensive or difficult to obtain, large datasets are rare. In the domain of medical image analysis, this is often the case. Although great efforts are being made to create a large publicly available dataset, generally speaking, the amount of available medical images and expert annotations are low. The cohort sizes can be extended by applying data-augmentation to the samples during the training (and test) phase. However, this is limited by the chosen transformations, which preferably insert a realistic prior in the transformed images. For example, if a model is trained to segment the left ventricle from computed tomography (CT) images and all patients have the same orientation, realistic rotations of the left ventricle (LV) can be inserted by rotating the images with some degrees. Flipping the images is a bad idea since this would never occur in this scenario. How much or which data-augmentation techniques need to be added is currently an art rather than a science. Moreover, some data-augmentation techniques are challenging to implement, for example, adding calcified areas in the aortic annular plane is challenging to achieve augmentation method because calcification can only appear inside the aortic annulus and not randomly in the image.

In this study, a tailored-data generation method was explored. This strategy might become an excellent candidate to generate data which meets specific criteria. The strategy was applied to images of the aortic annular plane (AAP). The generated images were evaluated, and we submitted an expert to a visual Turing test. The first

steps towards tailored image generation were explored.

## A.2 Material and methods

### A.2.1 MDCT images

The multidetector computed tomography (MDCT) images of 455 transcatheter aortic valve replacement (TAVR) patients with a certain degree of aortic stenosis were used during this study. The mean age of this cohort was $80.8 \pm 7.1$, and 55% of the patients were female. This dataset was the same as the study described in Section 5.2. The AAP was exported using the standard method [182]. The images of the AAP were used to train the models.

### A.2.2 Generative adversarial network

A generative adversarial network (GAN) uses two models: the generator model $\mathcal{G}$ and the discriminator model $\mathcal{D}$. The idea is that $\mathcal{G}$ tries to trick $\mathcal{D}$ by generating realistic images. The binary classification model $\mathcal{D}$, on the other hand, tries to classify generated images from $\mathcal{G}$ and the real dataset as generated or real [99].

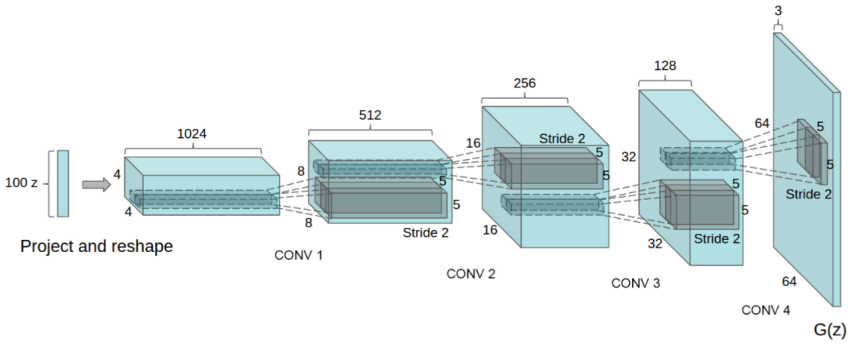



*Figure A.1: Overview of the architecture of the generator model $\mathcal{G}$. Latent space Z is a vector of size 100 in this example. Image from [235].*

Training GANs entails a couple of problems such as non-convergence, mode-collapse, diminished gradient, and unbalance between $\mathcal{G}$ and $\mathcal{D}$. Non-convergence occurs when the parameters of the models oscillate and never converge with poor performing models as a result. Mode-collapse is the phenomenon when the generator 'collapses' and is not able to produce the full variety of the training set. The diminished gradient is caused by a successful model $\mathcal{D}$ with a vanishing gradient in model $\mathcal{G}$ as a consequence. Finally, the unbalance between $\mathcal{D}$ and $\mathcal{G}$ can be caused by high sensitivity to the hyper-parameters and failing to reach a balance or Nash equilibrium [236]. GAN is based on zero-sum game or mini-max. Both models $\mathcal{D}$ and $\mathcal{G}$ want to

maximise their actions and minimise the opponent's actions. The Nash equilibrium is reached when the actions of an opponent do not influence one owns actions.

It has been shown that deep convolutional generative adversarial networks (DCGAN) can generate realistic images from a latent space Z [235]. In Figure A.1, an example of an architecture of a generator model is depicted.

### A.2.3 Training, evaluation and Z-arithmetic

The model and training scheme used in this study were inspired by [235]. After successful training, the generator model was able to generate realistic images from a latent space Z.

As a next step, the trained generator model $\mathcal{G}$ was used to generated images from random latent spaces. A virtual Turing test was performed where ten real and ten generated images were presented to an expert. At first, the expert was not informed about the experiment and was only asked to grade the image quality of the 20 aortic annular planes (with zero being the worse quality and five the best quality). In the second phase, the expert was informed about the experiment and was asked to label the generated and real images. Finally, we hypothesised that certain anatomical features in the generated images were linked to specific properties of the latent space Z. Therefore, exploratory arithmetic operations were performed on the latent spaces in order to measure the effect on specific features in the generated images.

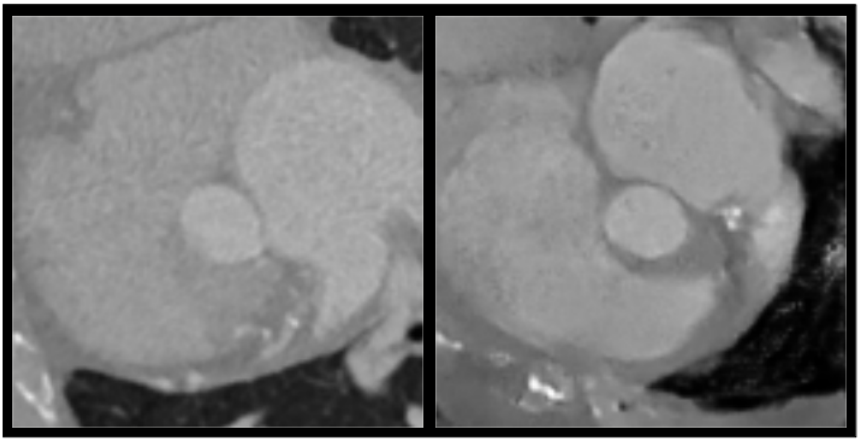


*Figure A.2: Comparison between a real (left) and generated (right) image of the aortic annular plane.*

## A.3 Results and discussion

The expert was presented with ten real and ten generated images. The mean quality score (in %) for the real images was 62% and for the generated images 48%. The difference may indicate that the generated images lacked some realistic details (Figure A.2). When asked to label the 20 images as real or generated, the expert

labelled 40% of the images correctly, which may indicate that the generated images were sufficiently realistic.

Arithmetic operations were performed on the latent space Z (Figure A.3). Anatomical shape differences of the left ventricle, lungs and others were observed. This observation could indicate that parts of the latent space may influence anatomical features.

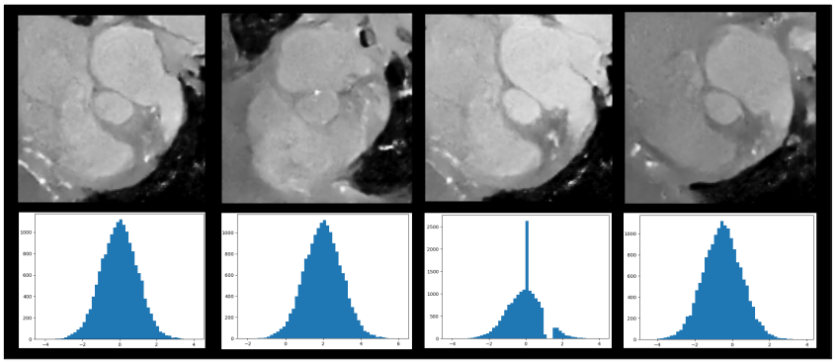


*Figure A.3: Generated images (top) and the histogram of their associated latent space Z (bottom).*

In conclusion, it was shown that DCGAN could be used to generate realistic medical images which can endure a visual Turing test. Experiments with the latent space indicate that generating images with specific criteria need further investigation.

# References

[99] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[182] A. M. Kasel, S. Cassese, S. Bleiziffer, M. Amaki, R. T. Hahn, A. Kastrati, and P. P. Sengupta, "Standardized imaging for aortic annular sizing: Implications for transcatheter valve selection", *JACC: Cardiovascular Imaging*, vol. 6, no. 2, pp. 249–262, 2013.

[235] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", *arXiv preprint arXiv: 1511.06434*, 2015.

[236] J. F. Nash *et al.*, "Equilibrium points in n-person games", *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

# B

# Light-weight algorithm for deep learning architecture evolution

Recent studies have shown that algorithms for evolving deep learning architectures for image classification can be used to generate high-performing deep learning models [230, 231]. These algorithms, however, require a lot of computation time and power. In this study, a light-weight algorithm for generating deep learning architectures for image-classification is proposed and validated.

## B.1 Introduction

Deep learning has become an established approach in the past decade for numerous image based-applications such as classification [23], segmentation [24], generation [99], and reinforcement learning [224]. All these applications share similar properties: a dataset, a well-formed cost function, a deep learning architecture and numerous enthusiastic training sessions. During these training sessions, the researcher(s) may start from an already established deep learning architecture and experiment with new ideas and some educated guesses. With each experiment or new idea, the deep learning architecture is manually adapted, trained and preferably compared to prior training sessions and already published data. Recent studies [230, 231] have proposed to abandon this manual step and automate this process by evolving architectures. One of the shared property of these studies is that they require a lot of computation time and power to obtain high-performing models.

In this study, a light-weight algorithm for generating deep learning architectures for image-classification is proposed and validated against the MNIST dataset [237]. It is shown that this method does not require substantial computation time and power

to obtain high-performing models.

## B.2 Material and methods

### B.2.1 Data

The MNIST dataset was used in this study to ensure readability and reproducibility. This dataset is used as a benchmark dataset in the supervised learning community and consists of grey-scale images of handwritten numbers and their associated label. The numbers in the images and labels range from 0 to 9 (Figure B.1). The dataset was split in a training dataset and a validation dataset to prevent data overlap during training and validation of the generated deep learning architectures.

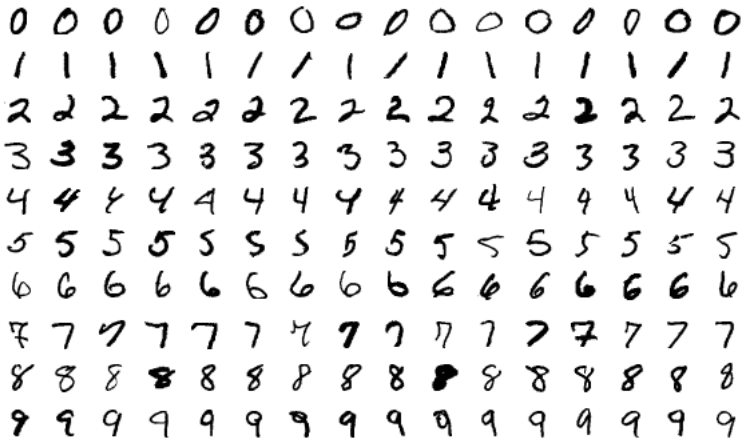


*Figure B.1: Examples from the MNIST dataset: handwritten numbers ranging from 0 to 9.*

### B.2.2 Individual

In this study, an evolution algorithm is applied to grow deep learning architectures. The individual in this process is defined to contain all elements that could mutate during this evolution. Elements such as the learning rate and a initially empty deep learning architecture are contained in each individual (Figure B.2). The deep learning architecture was represented by a directed graph which could grow, shrink or alter due to mutations which will be presented in the next section.

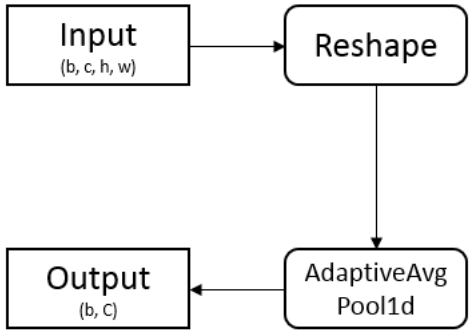



*Figure B.2: The initial deep learning architecture of an individual. The input consisted of samples with the following dimensions: a batch-size (b), number of channels (c), height (h) and width (w). The reshape operations flattened the number of channels, height and width and average pooling was applied over the input-signal. The final output consisted of the following dimensions: a batch-size (b) and the number of classes (C) of the classification task.*

Before the training, the graph needed to be resolved in order to enforce consistency. The first condition resolved potential size mismatches. The graph was traversed from the root until the end, and each node was checked. Subsequent nodes would receive the sizing information from the previous check. This step was necessary with operations that could alter the size of the input signal such as convolutions with stride $>1$, the number of features $>1$ or skip connections. The second condition made sure that the total size of the last node in the graph would not be smaller than the number of classes of the classification task. After successfully resolving the graph, the architecture was built and ready for training and validation.

### B.2.3 Mutations

Random mutations were required in order to evolve individuals. The process of trial-and-error during architecture development inspired the available mutations: altering the learning rate, adding or removing layers or altering existing layers. When the individual was selected for mutation, one mutation was selected at random and applied to the individual.

The following strategy was used for adding or removing layers and skip-connections [52]. In order to add a layer, a random node was selected from the graph. The size of this node was used to set (or alter) the size of the new layer. In order to remove a layer, a random node was selected from the graph and removed. In order to add a skip-connection, two random insertion nodes were selected from the graph. After sorting the nodes to ensure proper direction, a skip-connection between both nodes was added to the graph. Finally, to remove a skip-connection, a random skip-connection was selected and removed from the graph.

#### B.2.3.1 Mutation details

The following mutations were available in the evolution algorithm (inspired by [230]). All random alterations based on old values were obtained by multiplying the old value with a uniform random value between 0.5 and 1.5.

1. Altering the learning rate: generated at random based on the old value.
2. Insert a convolutional layer: with kernel size 3, padding 1, stride 1 and feature size equal to the number of channels or generated at random based on the old value.
3. Remove convolution layer.
4. Insert activation layer: insert a ReLU activation function.
5. Remove activation layer.
6. Insert batch normalisation layer.
7. Remove batch normalisation layer.
8. Keep training: the identity operation.
9. Add skip-connection.
10. Remove skip-connection.
11. Alter stride: randomly selecting a convolutional layer and alter the stride value (randomly selected from the following range: 1,2,4).
12. Alter number of features: randomly selecting a convolutional layer and alter the number of features based on the old value.

### B.2.4 Training and validation

The training set was used to train an individual's architecture for 25000 training steps to prevent overfitting and ensure fast evolution of individuals. The stochastic gradient descent with a momentum of 0.9 was used as optimiser with the negative log-likelihood as cost function. The learning rate was taken from the individual.

The entire validation set was used to validate the model, and a fitness score was set to the ratio of correct/incorrect number of classified images or accuracy of the model.

### B.2.5 Evolution algorithm

The following algorithm was used to evolve deep learning architectures for image classification. A population was trained with initially n empty individuals. After training, the individuals were validated. Next, a random selection was applied to the population where all alive individuals were randomly paired, and the individual with the largest fitness score was kept alive while the other one was marked as dead. The surviving individuals were randomly mutated, trained, validated and put back in the population. Finally, the process of random selection and mutation, training and validation were repeated for several generations.

## B.3 Results

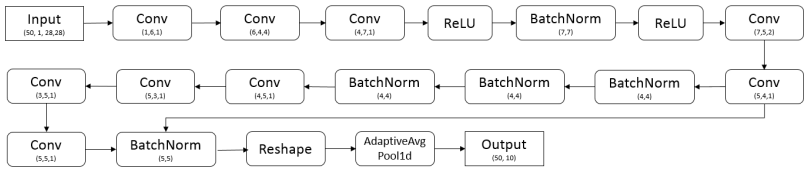



*Figure B.3: A grown deep learning architecture with a fitness score of 0.94. The convolutional nodes are marked with the following numbers (number of input features, number of output features, stride). The batch normalisation layers are marked with the number of input- and output features.*

A population of 10 individuals was used in the execution of our evolution algorithm. The initial individuals started with the lowest fitness value while mutated individuals progressed to higher fitness scores. After 100 generations, high fitness score reached a ceiling (Figure B.4). The entire evolution took 12 hours on a single graphical processing unit (GPU).

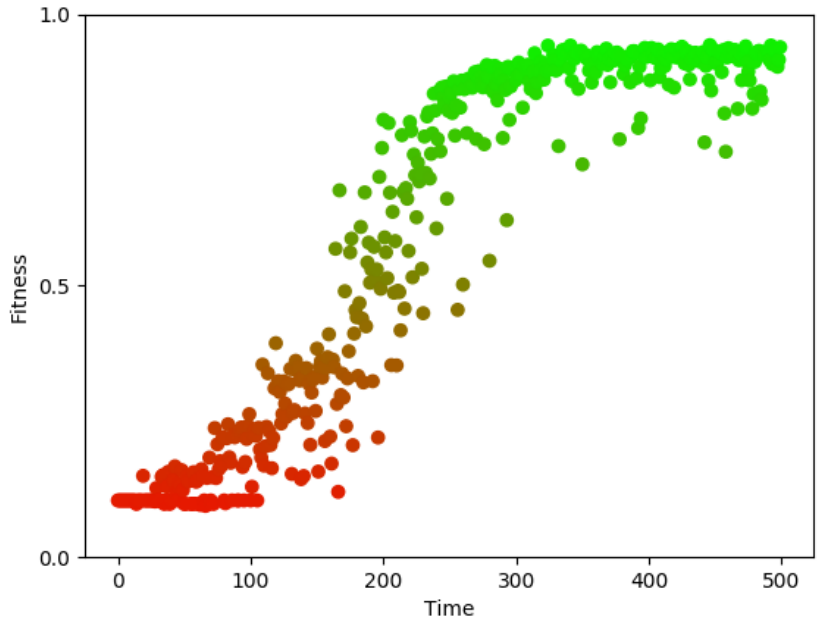



*Figure B.4: A plot of the fitness of the individuals in time (red=low fitness, green=high fitness value). Time is measured in amount of individuals.*

After 100 generations, an individual reached a fitness score of 0.94 in the population. Figure B.3 depicts the associated deep learning architecture of this individual. Note that there are only two ReLU activation functions in this architecture which is unusual

in deep learning architectures.

## B.4 Discussion

In this study, a light-weight algorithm for evolving deep learning architectures for image-classification is proposed. The algorithm is validated against the MNIST dataset, and after 500 mutations, an evolved deep learning model yielded a validation accuracy of 0.94.

In literature, high performing models have already been obtained with deep learning architecture evolution. In [231], the authors explored the idea of learning deep network structures automatically. The authors proposed a genetic algorithm to create new network structures. Tests on MNIST and CIFAR-10 have shown the methods ability to evolve and find high-quality structures. The complete evolution takes "2 GPU-days". In [230], the authors proposed a method that evolves network structures and applies simple evolutionary techniques at unprecedented scales to discover models for the CIFAR-10 and CIFAR-100 datasets. The complete evolution takes 256 hours wall time with 250 parallel workers. The previous studies clearly show that architecture evolution can be used to obtain high-performing deep learning models, however, on the condition that a high amount of computation time and power is available.

More research is required to see if other datasets such as CIFAR-10 and CIFAR-100 can be analysed with this method. It would also be interesting to see whether it is possible to grow novel blocks such as the residual or dense blocks.

## B.5 Conclusion

In this study, we showed that similar high performing deep learning models could be obtained without a high amount of computation time and power. Moreover, our evolution algorithm discovered an unusual deep learning architecture which used only two activation function.

## References

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

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

[99] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[224] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning", in *International conference on machine learning*, 2016, pp. 1928–1937.

[230] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers", in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, JMLR. org, 2017, pp. 2902–2911.

[231] L. Xie and A. Yuille, "Genetic cnn", in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1379–1388.

[237] Y. LeCun and C. Cortes, "MNIST handwritten digit database", 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/.

# C

# Echocardiographic analysis with deep learning: from cardiac cycle classification to cardiac output regression

## C.1 Introduction

The detection of the end-systolic (ES) and end-diastolic (ED) phase in ultrasound (US) series is an important step prior to medical image analysis and patient care. From these two phases, operators derive vital cardiac output, such as the ES- and ED volume.

In literature, the classification of the cardiac cycle from 3D US images has been performed with deep learning [238].

In this exploratory study, we measured the effect of medical image quantity in an automatic 4D US analysis workflow. A cycle classification model classified the frames of the varying length US series into three classes, and a regression model predicted the cardiac output.

## C.2 Material and methods

The cohort consisted of 4D echocardiography[1] of 128 patients. Each patient was analysed by an expert using 4D Cardio-View (TomTec, Unterschlessheim, Germany) which produced the necessary ground truth cardiac output (end systolic volume, systolic volume, end diastolic volume and ejection fraction). The length of the time series was not predetermined. (Figure C.1).

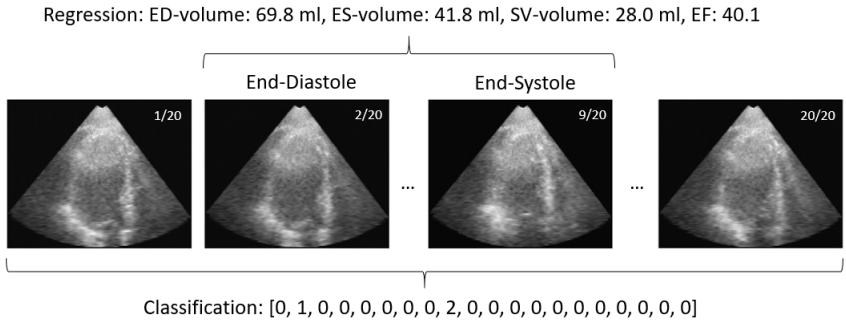



*Figure C.1: Example of US images from the cardiac cycle. The frames were classified as; others (0), ED (1) and ES (2). The regression task used the ED and ES images.*

### C.2.1 Classification

The cycle classification task was performed by a deep learning architecture consisting of an image encoder (CNN), a time series encoder (LSTM) and a funnel (Dense layers). A schematic overview of this architecture is depicted in Figure C.2. A novel referee function (inspired by [239]) was added to the weighted negative-log-likelihood cost function to enforce sureness and abolish doubt in the model. The referee function used the predicted classes and computed the absolute difference between the ground truth classes, thus pushing the model to certainty.

[1]Medical ultrasound is a diagnostic imaging technique to create an image of internal body structures. Ultrasound images are made by sending pulses of ultrasound (>20,000 Hz) into tissue using a probe and recording the different reflection properties.

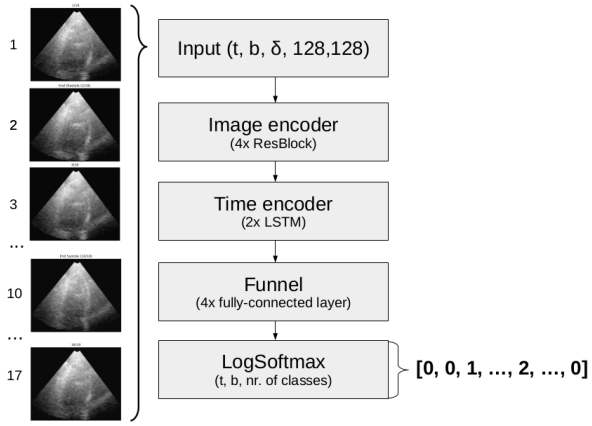



*Figure C.2: A schematic overview of the temporal classification architecture where t is the number of frames of the 4D image, b is the batch size, and δ is the number of images used in the configuration.*

### C.2.2 Regression

The regression model consisted of an image encoder and a funnel to predict the cardiac output. A schematic overview of this architecture is depicted in Figure C.3.

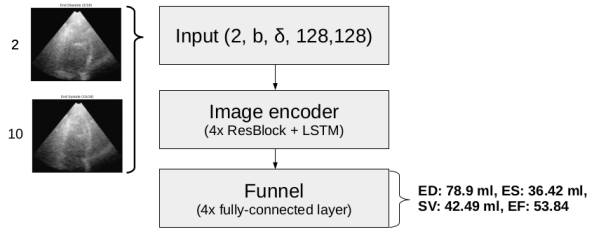



*Figure C.3: A schematic overview of the temporal regression architecture where b is the batch size, and δ is the number of images used in the configuration.*

### C.2.3 Configurations

In order to measure the effect of medical image quantity, we defined different configurations based on the number of images that were taken per axis: A=1, B=3, C=6 and D=13 (Table C.1).

## C.3 Results

All the models were cross-validated with k-folding (k=10). Both tasks yielded similar results for each configuration with a classification accuracy of 0.89, 0.88, 0.88 and 0.89 and a mean paired difference of 3.8, 4.2, 3.2 and 2.3 for configuration A, B,

| Config name | Selected slice indices | Total amount of images |
|---|---|---|
| A | 64 | 3 |
| B | 32, 64, 96 | 9 |
| C | 16, 32, 48, 64, 80, 96 | 18 |
| D | 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96, 104 | 39 |

*Table C.1: Image quantity configuration overview. The selected slice indices refer to the three axes: sagittal, coronal and axial.*

C and D (Figure C.4). The training time for configuration A was substantially faster compared to configuration D .

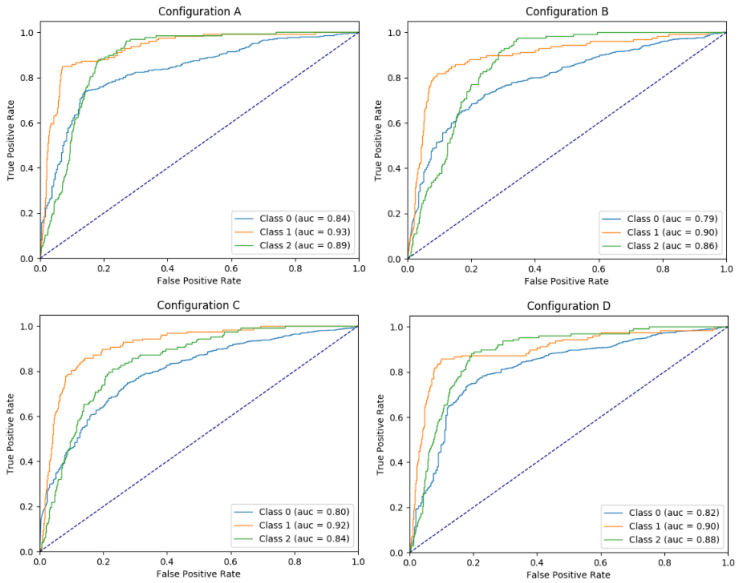



*Figure C.4: The receiver operating characteristic curves of the configurations.*

## C.4 Discussion and conclusion

These findings require further investigation; however, since the advent of AI, the amount of available medical data has been increasing. Subsequently, the question arises how much data is required for useful insights?

## References

[238] A. MeidellFiorito, A. Østvik, E. Smistad, S. Leclerc, O. Bernard, and L. Lovstakken, "Detection of Cardiac Events in Echocardiography Using 3D convolutional Recurrent Neural Networks", in *2018 IEEE International Ultrasonics Symposium (IUS)*, IEEE, 2018, pp. 1–4.

[239] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection", in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

# D

# Cohort overview

## D.1 Aortic cohort

The anonymised cohort consisted of volumetric MDCT images, which were acquired to support the pre-operative phase of a TAVR procedure. Most of the MDCT images were contrast-enhanced and contained a certain degree of aortic stenosis. The cohort consisted of 437 tricuspid patients 36 bicuspid patients and the average row and column size were 512.1, 511.7. Average slice thickness of the images was 0.8 mm, and the average pixel spacing was 0.53 mm for both row- and column spacing.

As introduced in chapter 3, all DICOM images contain meta-information which can be extracted and provide insights into the cohort's demography. In the following sections, the next topics; gender, age, contrast and manufacturer are summarised.

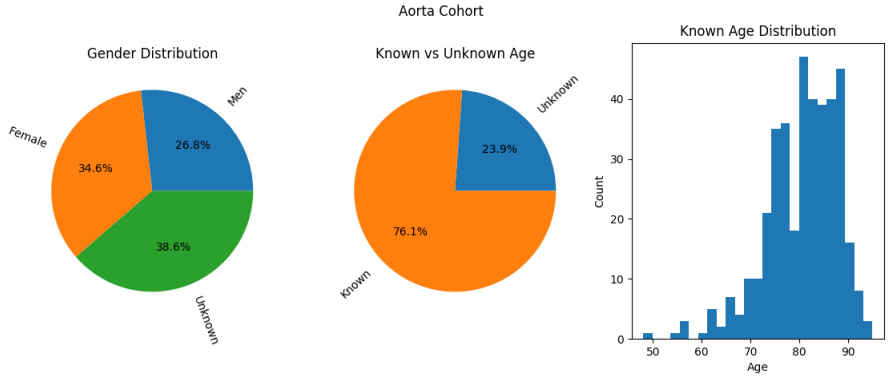



*Figure D.1: Gender and age distribution of the aortic cohort*

The gender was known in almost 60% of the patients, and the age was known in 76% of the patients. The lacking gender or age information indicates that the meta-data is incomplete or has been removed. The known age distribution had a median value of 81 and lower and upper quartile values of 76 and 86. The minimum and maximum age were 48 and 95 (Figure D.1).

Almost 90% of the used contrast information in the aortic cohort was known. These known contrasts were filtered and categorised into 11 categories. Note that 32% of the known contrasts was called "Applied", which does not provide information about which type of contrast was used (Figure D.2).

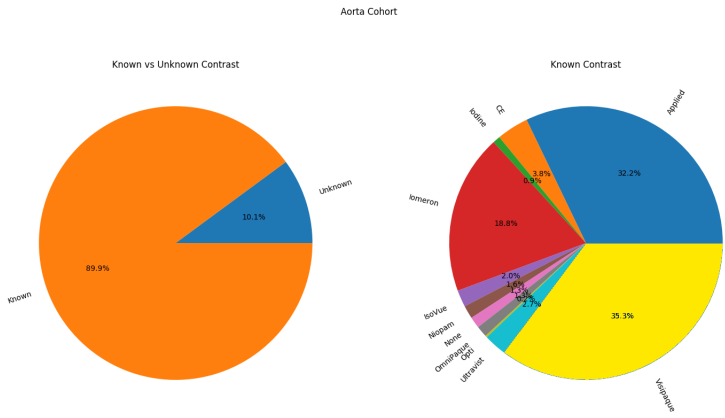



*Figure D.2: Contrast information of the aortic cohort*

Unfortunately, more than 90% of the manufacturers of the recording devices were unknown. The eight unique manufacturers were Somatom, Discovery, Brilliance, Revolution CT, Sensation Cardiac 64, LightSpeed, iCT and Aquilion (Figure D.3).

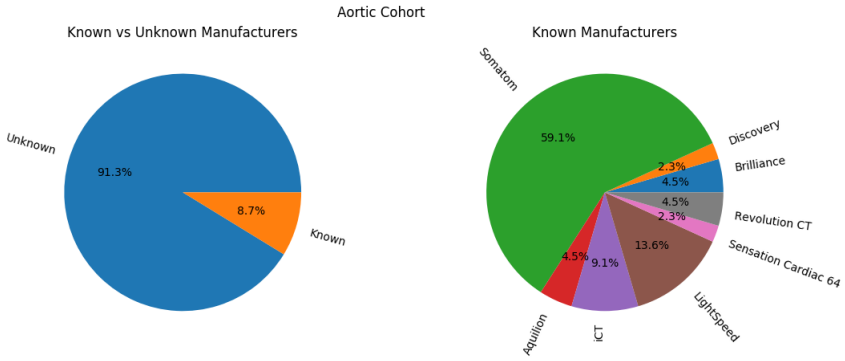



*Figure D.3: Manufacturers information of the aortic cohort*

### D.1.1 Classification cohort

In this dataset, a subset of patients from the aortic cohort was taken and included 48 tricuspid patients and 48 bicuspid patients. The median age was 82 with lower and upper quartile of 76 and 86. The minimum and maximum age were 63 and 91. The unique contrasts used in this dataset were; Niopam, Visipaque, IsoVue, Iomeron and Ultravist. The unique manufacturer's names of the recording devices were; Discovery, Somatom and Revolution.

### D.1.2 Membranous Septum Detection

A subset of the aortic cohort was used in this dataset due to the availability of the ground truth. The trained operators detected the membranous septum of 278 patients.

## D.2 Mitral cohort

This anonymised cohort of 71 patients was collected from multiple centra and is summarised in this final section.

The gender was known in 80% of the patients, and the age was known in 77% of the patients. The known age distribution had a median value of 77 and lower and upper quartile values of 71 and 82. The minimum and maximum age were 0 and 90. The minimum age indicates an error in the available meta-data of the DICOM images (Figure D.4).

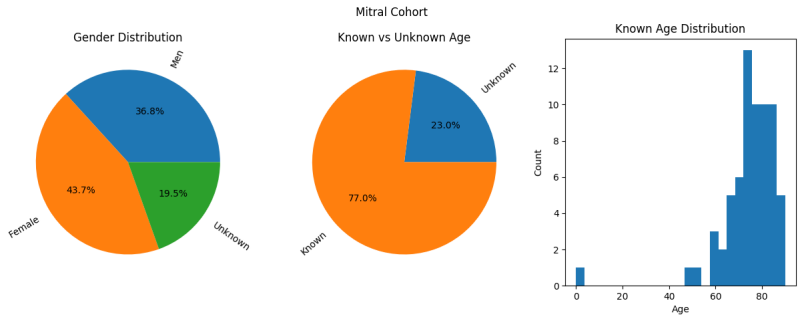



*Figure D.4: Gender and age distribution of the mitral cohort*

Almost 90% of the used contrast in the mitral cohort were known. These known contrasts were filtered and categorised into eight categories. Note that 32% of the known contrasts was called "Applied", which does not provide information about which type of contrast was used (Figure D.5).

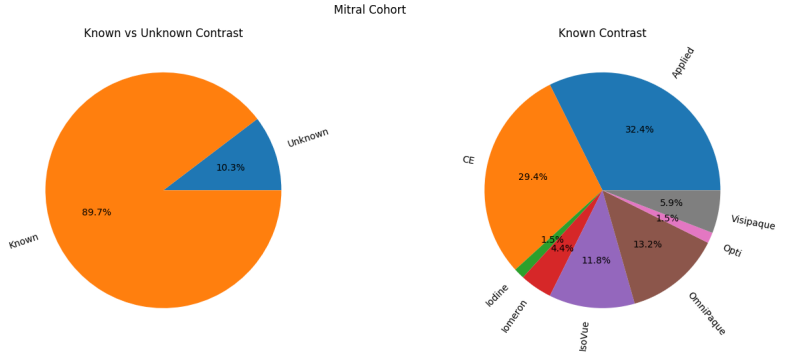



*Figure D.5: Contrast information of the mitral cohort*

Almost 60% of the manufacturers of the recording devices were unknown. The five unique manufacturers were Somatom, Discovery, Revolution CT, iCT and Aquilion (Figure D.6).

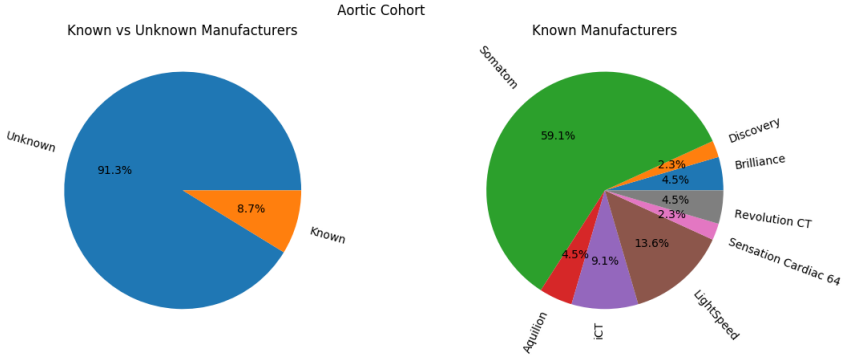



*Figure D.6: Manufacturers information of the mitral cohort*